An Introduction to Linear Algebra

Andrew D. Hwang

February 2015

To the Student

Linear algebra comprises a variety of topics and viewpoints, including computational machinery (matrices), abstract objects (vector spaces), mappings (linear transformations), and the "fine structure" of linear transformations (diagonalizability).

Like all mathematical subjects at an introductory level, linear algebra is driven by examples and comprehended by unfamiliar theory. Each of the preceding topics can be difficult to assimilate until the others are understood. You, the reader, consequently face a chicken-and-egg problem: Examples appear unconnected without a theoretical framework, but theory without examples tends to be dry and unmotivated.

This preface sketches an overview of the entire book by looking at a family of representative examples. The "universe" is the Cartesian plane $\mathbf{R}^2$, whose coordinates we denote (x^1, x^2). (The use of indices instead of the more familiar (x, y) will economize our use of letters, particularly when we begin to study functions of arbitrarily many variables. The use of superscripts as *indices* rather than as exponents highlights important, subtle structure in formulas.) We view ordered pairs as individual entities, and write $\boldsymbol{x} = (x^1, x^2)$.

Vectors and Vector Spaces

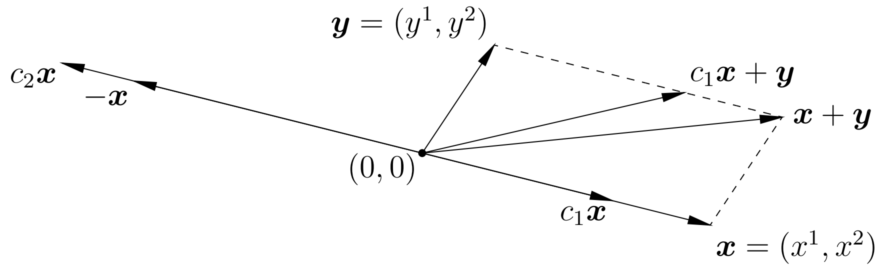
We view an ordered pair $\boldsymbol{x}$ as a *vector*, a type of object that can be added to another vector, or multiplied by a real constant (called a *scalar*) to obtain another vector. If $\boldsymbol{x} = (x^1, x^2)$ and $\boldsymbol{y} = (y^1, y^2)$, and if c is a scalar, we define

$$\boldsymbol{x} + \boldsymbol{y} = (x^1 + y^1, x^2 + y^2), \quad c\boldsymbol{x} = (cx^1, cx^2).$$

The set $\mathbf{R}^2$ equipped with these operations is said to be a *vector space*.

The vector $\boldsymbol{x} = (x^1, x^2)$ in the plane may be viewed geometrically as the arrow with its tail at the origin $\mathbf{0} = (0, 0)$ and its tip at the

point $\boldsymbol{x}$. Vector addition corresponds to forming the parallelogram with sides $\boldsymbol{x}$ and $\boldsymbol{y}$, and taking $\boldsymbol{x}+\boldsymbol{y}$ to be the far corner. Scalar multiplication $c\boldsymbol{x}$ corresponds to "stretching" $\boldsymbol{x}$ by a factor of c if $c>0$, or to stretching by a factor of $|c|$ and reversing the direction if $c<0$.



Linear Transformations

In linear algebra, most mappings send vector spaces to vector spaces. The special properties dictated by linear algebra may be written

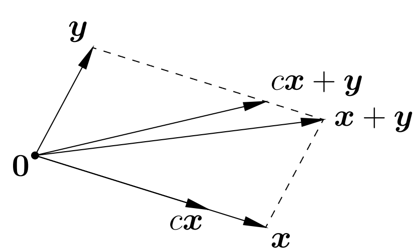
$$\left.\begin{aligned} S(\boldsymbol{x}+\boldsymbol{y}) &= S(\boldsymbol{x})+S(\boldsymbol{y}),\\ S(c\boldsymbol{x}) &= cS(\boldsymbol{x}) \end{aligned}\right\} \qquad \text{for all vectors } \boldsymbol{x},\ \boldsymbol{y}, \text{ all scalars } c.$$

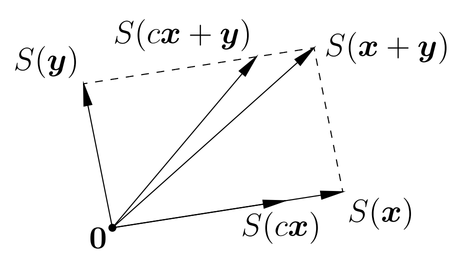
For technical convenience, these conditions are often expressed as a single condition

$$S(c\boldsymbol{x}+\boldsymbol{y}) = cS(\boldsymbol{x})+S(\boldsymbol{y}) \qquad \text{for all vectors } \boldsymbol{x},\ \boldsymbol{y}, \text{ all scalars } c.$$

A mapping S satisfying this conditions is called a *linear transformation*.

Geometrically, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are arbitrary vectors and c is a scalar, so that $c\boldsymbol{x}+\boldsymbol{y}$ is the far corner of the parallelogram with sides $c\boldsymbol{x}$ and $\boldsymbol{y}$, then $S(c\boldsymbol{x}+\boldsymbol{y}) = cS(\boldsymbol{x})+S(\boldsymbol{y})$ is the far corner of the parallelogram with sides $S(c\boldsymbol{x}) = cS(\boldsymbol{x})$ and $S(\boldsymbol{y})$. A linear transformation S therefore maps an arbitrary parallelogram to a parallelogram in an obvious (and restrictive) sense.





The special vectors $\boldsymbol{e}_1 = (1, 0)$ and $\boldsymbol{e}_2 = (0, 1)$ are the *standard basis* of $\mathbf{R}^2$. Every vector in $\mathbf{R}^2$ can be expressed uniquely as a *linear combination*:

$$\begin{aligned}\boldsymbol{x} = (x^1, x^2) &= (x^1, 0) + (0, x^2)\\ &= x^1(1, 0) + x^2(0, 1)\\ &= x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2.\end{aligned}$$

Formally, a linear transformation "distributes" over an arbitrary linear combination. In detail, if $S : \mathbf{R}^2 \to \mathbf{R}^2$ is a linear transformation, repeated application of the defining properties gives

$$\begin{aligned}S(\boldsymbol{x}) &= S(x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2)\\ &= S(x^1\boldsymbol{e}_1) + S(x^2\boldsymbol{e}_2)\\ &= x^1S(\boldsymbol{e}_1) + x^2S(\boldsymbol{e}_2).\end{aligned}$$

This innocuous equation expresses a remarkable conclusion: *A linear transformation $S : \mathbf{R}^2 \to \mathbf{R}^2$ is completely determined by its values $S(\boldsymbol{e}_1)$, $S(\boldsymbol{e}_2)$ at two vectors.*

Matrix Representation

To study vector spaces and linear transformations in greater detail, we will represent vectors and linear transformations as rectangular arrays of numbers, called *matrices*. The first chapter of the book introduces matrix notation, a central piece of computational machinery in linear algebra. Here we focus on motivation and geometric intuition, using the special case of linear transformations from the plane to the plane.

We use the notational convention in which vectors are written as *columns*:

$$\boldsymbol{x} = (x^1, x^2) = \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}.$$

With this notation, a linear transformation $S : \mathbf{R}^2 \to \mathbf{R}^2$ is completely and uniquely specified by scalars A_i^j satisfying

$$S(\boldsymbol{e}_1) = A_1^1\boldsymbol{e}_1 + A_1^2\boldsymbol{e}_2 = \begin{bmatrix} A_1^1 \\ A_1^2 \end{bmatrix}, \quad S(\boldsymbol{e}_2) = A_2^1\boldsymbol{e}_1 + A_2^2\boldsymbol{e}_2 = \begin{bmatrix} A_2^1 \\ A_2^2 \end{bmatrix}.$$

The (*standard*) *matrix* of S assembles these into a rectangular array A:

$$A = \begin{bmatrix} S(\boldsymbol{e}_1) & S(\boldsymbol{e}_2) \end{bmatrix} = \begin{bmatrix} A_1^1 & A_2^1 \\ A_1^2 & A_2^2 \end{bmatrix}.$$

The real number A^i_j in the ith row and jth column of A is called the (i, j)-*entry*, and encodes the dependence of the ith output variable on the jth input variable.

Matrix Multiplication

For all $\boldsymbol{x}$, we have

$$\begin{aligned} S(\boldsymbol{x}) &= x^1 S(\boldsymbol{e}_1) + x^2 S(\boldsymbol{e}_2) \\ &= x^1 \begin{bmatrix} A^1_1 \\ A^2_1 \end{bmatrix} + x^2 \begin{bmatrix} A^1_2 \\ A^2_2 \end{bmatrix} = \begin{bmatrix} A^1_1 x^1 + A^1_2 x^2 \\ A^2_1 x^1 + A^2_2 x^2 \end{bmatrix}. \end{aligned}$$

The expression on the right may be interpreted as a type of "product" of the matrix of S and the column vector of $\boldsymbol{x}$:

$$\begin{bmatrix} y^1 \\ y^2 \end{bmatrix} = \begin{bmatrix} A^1_1 x^1 + A^1_2 x^2 \\ A^2_1 x^1 + A^2_2 x^2 \end{bmatrix} = \begin{bmatrix} A^1_1 & A^1_2 \\ A^2_1 & A^2_2 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix},$$

or simply $\boldsymbol{y} = A\boldsymbol{x}$. The second equality *defines* the product of the "2×2 square matrix A" and the "2×1 column matrix $\boldsymbol{x}$".

Particularly when the number of variables is large, sigma (summation) notation comes into its own, both condensing common expressions and highlighting their structure. The relationship between the inputs x^j and the outputs y^i of a linear transformation may be written $y^i = \sum_j A^i_j x^j$.*

Composition of Linear Transformations

If $T : \mathbf{R}^2 \to \mathbf{R}^2$ is a linear transformation with matrix

$$B = \begin{bmatrix} B^1_1 & B^1_2 \\ B^2_1 & B^2_2 \end{bmatrix},$$

the *composition* $T \circ S : \mathbf{R}^2 \to \mathbf{R}^2$, defined by $(T \circ S)(\boldsymbol{x}) = T\big(S(\boldsymbol{x})\big)$, is easily shown to be a linear transformation (check this). The associated

*Physicists sometimes go further, omitting the summation sign and implicitly summing over every index that appears both as a subscript and as a superscript: $y^i = A^i_j x^j$. This book does not use this "Einstein summation convention", but the *possibility of doing so* explains our use of superscripts as indices, despite the potential risk of reading superscripts as exponents. Exponents appear so rarely in linear algebra that mentioning them at each occurrence is feasible.

matrix is a "product" of the matrices B and A of the transformations T and S. To determine the entries of this product, note that by definition,

$$\begin{aligned} T(\boldsymbol{e}_1) &= B_1^1\boldsymbol{e}_1 + B_1^2\boldsymbol{e}_2, & S(\boldsymbol{e}_1) &= A_1^1\boldsymbol{e}_1 + A_1^2\boldsymbol{e}_2, \\ T(\boldsymbol{e}_2) &= B_2^1\boldsymbol{e}_1 + B_2^2\boldsymbol{e}_2, & S(\boldsymbol{e}_2) &= A_2^1\boldsymbol{e}_1 + A_2^2\boldsymbol{e}_2. \end{aligned}$$

Consequently,

$$\begin{aligned} TS(\boldsymbol{e}_1) &= T(A_1^1\boldsymbol{e}_1 + A_1^2\boldsymbol{e}_2) \\ &= A_1^1 T(\boldsymbol{e}_1) + A_1^2 T(\boldsymbol{e}_2) \\ &= A_1^1(B_1^1\boldsymbol{e}_1 + B_1^2\boldsymbol{e}_2) + A_1^2(B_2^1\boldsymbol{e}_1 + B_2^2\boldsymbol{e}_2) \\ &= (B_1^1A_1^1 + B_2^1A_1^2)\boldsymbol{e}_1 + (B_1^2A_1^1 + B_2^2A_1^2)\boldsymbol{e}_2; \end{aligned}$$

similarly (check this),

$$TS(\boldsymbol{e}_2) = (B_1^1A_2^1 + B_2^1A_2^2)\boldsymbol{e}_1 + (B_1^2A_2^1 + B_2^2A_2^2)\boldsymbol{e}_2.$$

Since the coefficients of $T(\boldsymbol{e}_1)$ give the first column of the matrix of TS and the coefficients of $T(\boldsymbol{e}_2)$ give the second column of the matrix, we are led to define the *matrix product* by

$$BA = \begin{bmatrix} B_1^1 & B_2^1 \\ B_1^2 & B_2^2 \end{bmatrix} \begin{bmatrix} A_1^1 & A_2^1 \\ A_1^2 & A_2^2 \end{bmatrix} = \begin{bmatrix} B_1^1A_1^1 + B_2^1A_1^2 & B_1^1A_2^1 + B_2^1A_2^2 \\ B_1^2A_1^1 + B_2^2A_1^2 & B_1^2A_2^1 + B_2^2A_2^2 \end{bmatrix}.$$

This forbidding collection of formulas is clarified by summation notation:

$$(BA)_j^i = \sum\nolimits_k B_k^i A_j^k.$$

The preceding equation has precisely the same form for matrices of arbitrary size, and furnishes our general definition of matrix multiplication. (Convince yourself that the entries of BA are given by the preceding summation formula.)

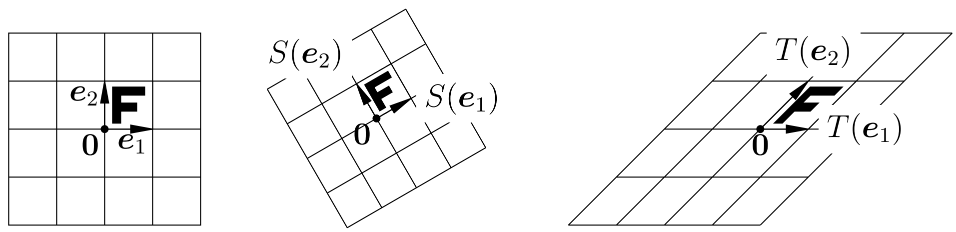
When working computationally with specific matrices, the *formula* is generally less important than the *procedure* encoded by the formula. First, define the "product" of a "row" and a "column" by

$$\begin{bmatrix} b & b' \end{bmatrix} \begin{bmatrix} a \\ a' \end{bmatrix} = \begin{bmatrix} ba + b'a' \end{bmatrix}.$$

Now, to find the (i, j)-entry of the product BA, multiply the ith row of B by the jth column of A. For example, to find the entry in the first row and second column of BA, multiply the first row of B by the second column of A.

Geometry of Linear Transformations

Consider the linear transformation S that rotates $\mathbf{R}^2$ about the origin by $\frac{\pi}{6}$, and T that shears the plane horizontally by one unit:



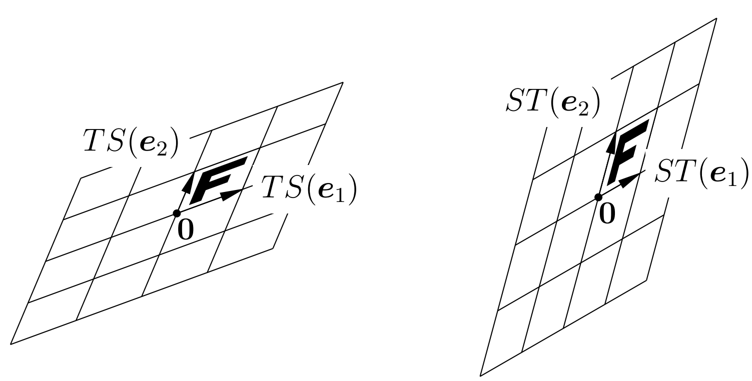
The matrix of each may be read off the images of the standard basis vectors. Thus,

$$
\begin{aligned}
A &= \begin{bmatrix} S(\boldsymbol{e}_1) & S(\boldsymbol{e}_2) \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{6} & \cos\frac{2\pi}{3} \\ \sin\frac{\pi}{6} & \sin\frac{2\pi}{3} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}, \\
B &= \begin{bmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.
\end{aligned}
$$

The composite transformations TS (rotate, then shear) and ST (shear, then rotate) are linear, and their matrices may be found by matrix multiplication:

$$
BA = \frac{1}{2}\begin{bmatrix} \sqrt{3}+1 & \sqrt{3}-1 \\ 1 & \sqrt{3} \end{bmatrix}, \qquad AB = \frac{1}{2}\begin{bmatrix} \sqrt{3} & \sqrt{3}-1 \\ 1 & \sqrt{3}+1 \end{bmatrix}.
$$

Note carefully that $TS \neq ST$. Composition of linear transformations, and consequently multiplication of square matrices, is generally a non-commutative operation.



Diagonalization

The so-called *identity matrix*

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

corresponds to the identity mapping $I(\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x}$ in $\mathbf{R}^2$. Generally, if λ^1 and λ^2 are real numbers, the *diagonal matrix*

$$\begin{bmatrix} \lambda^1 & 0 \\ 0 & \lambda^2 \end{bmatrix}$$

corresponds to *axial scaling*, $(x^1, x^2) \mapsto (\lambda^1 x^1, \lambda^2 x^2)$. Diagonal matrices are among the simplest matrices. In particular, if n is a positive integer, the nth power of a diagonal matrix is trivially calculated:

$$\begin{bmatrix} \lambda^1 & 0 \\ 0 & \lambda^2 \end{bmatrix}^n = \begin{bmatrix} (\lambda^1)^n & 0 \\ 0 & (\lambda^2)^n \end{bmatrix}.$$

The solution to a variety of mathematical problems rests on our ability to compute arbitrary powers of a matrix. We are naturally led to ask: If $S : \mathbf{R}^2 \to \mathbf{R}^2$ is a linear transformation, does there exist a coordinate system in which S acts by axial scaling? This question turns out to reduce to existence of scalars λ^1 and λ^2, and of non-zero vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, such that

$$S(\boldsymbol{v}_1) = \lambda^1 \boldsymbol{v}_1, \qquad S(\boldsymbol{v}_2) = \lambda^2 \boldsymbol{v}_2.$$

Each λ^i is an *eigenvalue* of S; each $\boldsymbol{v}_i$ is an *eigenvector* of S. A pair of non-proportional eigenvectors in the plane is an *eigenbasis* for S.

A linear transformation may or may not admit an eigenbasis. The rotation S of the preceding example has no real eigenvalues at all. The shear T has one real eigenvalue, and admits an eigenvector, but has no eigenbasis. The compositions TS and ST both turn out to admit eigenbases.

Structural Summary

Basic linear algebra has three parallel "levels". In increasing order of abstraction, they are:

(i) "The level of entries": Column vectors and matrices written out as arrays of numbers. ($y^i = \sum_j A^i_j x^j$.)

(ii) "The level of matrices": Column vectors and matrices written as single entities. ($\boldsymbol{y} = A\boldsymbol{x}$.)

(iii) "The abstract level": Vectors (defined axiomatically) and linear transformations (mappings that distribute over linear combinations). ($\boldsymbol{y} = S(\boldsymbol{x})$.)

Linear algebra is, at heart, the study of linear combinations and mappings that "respect" them. Along your journey through the material, strive to detect the levels' respective viewpoints and idioms. Among the most universal idioms is this: *A linear combination of linear combinations is a linear combination.*

Matrices are designed expressly to handle the bookkeeping details. In summation notation at level (i), if

$$z^i = \sum_{k=1}^{m} B^i_k y^k \quad \text{and} \quad y^k = \sum_{j=1}^{n} A^k_j x^j,$$

then substitution of the second into the first gives

$$z^i = \sum_{k=1}^{m} B^i_k \sum_{j=1}^{n} A^k_j x^j = \sum_{j=1}^{n} \left(\sum_{k=1}^{m} B^i_k A^k_j \right) x^j = \sum_{j=1}^{n} (BA)^i_j x^j.$$

Once we establish properties of matrix operations, the preceding can be distilled down to an extremely simple computation at level (ii): If $\boldsymbol{z} = B\boldsymbol{y}$ and $\boldsymbol{y} = A\boldsymbol{x}$, then $\boldsymbol{z} = B(A\boldsymbol{x}) = (BA)\boldsymbol{x}$.

Organization of the Book

The first chapter introduces real matrices as formally and quickly as feasible. The goal is to construct machinery for flexible computation. The book proceeds to introduce vector spaces and their properties, two auxiliary pieces of algebraic machinery (the dot product, and the determinant function on square matrices, each of which has a useful geometric interpretation), linear transformations and their properties, and diagonalization.

Of necessity, the motivation for a particular definition may not be immediately apparent. At each stage, we are merely generalizing and systematizing the preceding summary. It may help to review this preface periodically as you proceed through the material.

Contents

Chapter 1

Matrix Algebra

Definition 1.1. Let m and n be positive integers. An $m \times n$ *matrix* is a rectangular array having m rows and n columns:

$$\begin{bmatrix} A_1^1 & A_2^1 & \dots & A_n^1 \\ A_1^2 & A_2^2 & \dots & A_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ A_1^m & A_2^m & \dots & A_n^m \end{bmatrix},$$

or simply $[A_j^i]$ for brevity if the size is implicit (or unimportant). The (i, j) *entry* A_j^i is written in the ith row and jth column.

If $A = [A_j^i]$ and $B = [B_j^i]$ are $m \times n$ real matrices, their *sum* is the $m \times n$ matrix defined by $A + B = [A_j^i + B_j^i]$.

If c is real, we define *scalar multiplication* by $cA = c[A_j^i] = [cA_j^i]$.

Remark 1.2. The superscripts are *row indices*, not exponents. Explicit mention is made in this book when superscripts do stand for exponents. The idiomatic value of this notational convention will gradually become apparent.

When we write $[A_j^i]$, the indices i and j are *dummies*, having no meaning outside the brackets. We may use any convenient letters without changing the meaning, e.g., $[A_j^i] = [A_\ell^k]$, so long as the row index runs from 1 to m and the column index runs from 1 to n. What matters is that an unambiguous matrix entry is associated to each pair of indices.

Remark 1.3. In this book, the entries of a matrix are usually real numbers. We may speak of *real matrices* to emphasize this fact. The set of all $m \times n$ real matrices is denoted $\mathbf{R}^{m \times n}$.

However, the entries of a matrix might be other kinds of numbers (leading us to speak, for example, of *integer matrices* or *complex matrices*), or even other matrices (in which case we speak of *block matrices*). The space of all complex $m \times n$ matrices is denoted $\mathbf{C}^{m\times n}$.

Definition 1.4. A matrix of size $n \times 1$ is called a *column*. We denote the set of columns by $\mathbf{R}^n$ (read "RN") rather than by $\mathbf{R}^{n\times 1}$.

A matrix of size $1 \times n$ is called a *row*. We denote the set of rows by $(\mathbf{R}^n)^*$ (read "RN star") rather than by $\mathbf{R}^{1\times n}$.

If $A = [A^i_j]$ is an $m \times n$ matrix, A^i denotes the ith row, an element of $(\mathbf{R}^n)^*$, and A_j denotes the jth column, in $\mathbf{R}^m$.

Definition 1.5. If i and j are integers, the expression

$$\delta^i_j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

is called the *Kronecker delta-symbol*.

Example 1.6. The following matrices are all defined by $A^i_j = \delta^i_j$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Definition 1.7. Let n be a positive integer. For each $j = 1, \dots, n$, we define $\boldsymbol{e}_j$ to be the element of $\mathbf{R}^n$ whose ith row is δ^i_j. That is, the jth row is 1 and all other entries are 0:

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \; \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

The ordered set $(\boldsymbol{e}_j)_{j=1}^n$ is called the *standard basis* of $\mathbf{R}^n$.

Definition 1.8. Let n be a positive integer. For each $i = 1, \dots, n$, we define $\boldsymbol{e}^i$ to be the element of $(\mathbf{R}^n)^*$ whose jth column is δ^i_j. That is, the ith column is 1 and all other entries are 0:

$$\boldsymbol{e}^1 = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}, \quad \boldsymbol{e}^2 = \begin{bmatrix} 0 & 1 & \dots & 0 \end{bmatrix}, \; \dots, \; \boldsymbol{e}^n = \begin{bmatrix} 0 & 0 & \dots & 1 \end{bmatrix}.$$

The ordered set $(\boldsymbol{e}^i)_{i=1}^n$ is called the *standard dual basis* of $(\mathbf{R}^n)^*$.

Definition 1.9. Let n be a positive integer. A matrix of size $n \times n$ is called a *square matrix*.

The *main diagonal* of a square matrix is the set of entries $(A_j^j)_{j=1}^n$ running from the upper left to the lower right. A square matrix is *diagonal* if every non-diagonal entry is 0. The diagonal matrix I_n in $\mathbf{R}^{n\times n}$ defined by $A_j^i = \delta_j^i$ is called the $(n \times n)$ *identity matrix*.

Example 1.10. The identity matrices are $I_1 = \begin{bmatrix}1\end{bmatrix}$, $I_2 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$,

$$I_3 = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, \qquad I_4 = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}, \qquad I_n = \begin{bmatrix}1 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 1\end{bmatrix}.$$

Remark 1.11. The jth column of I_n is the standard basis vector $\boldsymbol{e}_j$. In other words, I_n may be written as a (block) row matrix whose jth column is $\boldsymbol{e}_j$, namely, $I_n = [\boldsymbol{e}_1 \quad \boldsymbol{e}_2 \quad \dots \quad \boldsymbol{e}_n]$.

Similarly, the ith row of I_n is $\boldsymbol{e}^i$, so I_n may be written as a (block) column matrix whose ith row is $\boldsymbol{e}^i$.

Definition 1.12. If $A = [A_j^i] \in \mathbf{R}^{m\times n}$, the *transpose* of A is the matrix A^T in $\mathbf{R}^{n\times m}$ whose (i, j) entry is A_i^j.

Example 1.13. $\begin{bmatrix}a^1 & a^2 & a^3\\ b^1 & b^2 & b^3\end{bmatrix}^\mathsf{T} = \begin{bmatrix}a^1 & b^1\\ a^2 & b^2\\ a^3 & b^3\end{bmatrix}$, and *vice versa*.

Remark 1.14. The transpose of a row matrix is a column, and *vice versa*. Particularly, the transpose defines a map from $\mathbf{R}^n$ (the set of columns) to $(\mathbf{R}^n)^*$ (the set of rows). For example, $\boldsymbol{e}_i^\mathsf{T} = \boldsymbol{e}^i$ and $(\boldsymbol{e}^i)^\mathsf{T} = \boldsymbol{e}_i$.

Generally, the ith row of A^T is the transpose of the ith column of A, and the jth column of A^T is the transpose of the jth row of A. The transpose of A^T is A itself: $(A^\mathsf{T})^\mathsf{T} = A$.

Remark 1.15. If $A = [A_j^i]$ and $B = [B_j^i]$ are in $\mathbf{R}^{m\times n}$, and if c is real, then $(cA + B)^\mathsf{T} = cA^\mathsf{T} + B^\mathsf{T}$. This is immediate by comparing entries: $(cA + B)_i^j = cA_i^j + B_i^j$.

1.1 Matrix Multiplication

An $m \times n$ real matrix is a "data structure" containing mn real numbers. Matrix addition and scalar multiplication merely "parallelize"

real addition and real multiplication. A third operation, matrix multiplication, "mixes" entries. The definition is introduced in two stages.

Definition 1.16. Let m, p, and n be positive integers. Assume

$$\boldsymbol{\ell} = [\ell_k] = [\ell_1 \quad \ell_2 \quad \dots \quad \ell_p] \in (\mathbf{R}^p)^*, \qquad \boldsymbol{v} = [v^k] = \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ v^p \end{bmatrix} \in \mathbf{R}^p.$$

We define the *product* or *dual pairing* $\langle \mid \rangle_{\mathbf{R}^p} : (\mathbf{R}^p)^* \times \mathbf{R}^p \to \mathbf{R}$ by

$$\langle \boldsymbol{\ell} \mid \boldsymbol{v} \rangle_{\mathbf{R}^p} = [\ell_1 \ \ell_2 \ \dots \ \ell_p] \begin{bmatrix} v^1 \\ v^2 \\ \vdots \\ v^p \end{bmatrix} = \ell_1 v^1 + \ell_2 v^2 + \cdots + \ell_p v^p = \sum_{k=1}^{p} \ell_k v^k.$$

Generally, if $A = [A^i_k] \in \mathbf{R}^{m \times p}$ and $B = [B^k_j] \in \mathbf{R}^{p \times n}$, their (matrix) *product* is the $m \times n$ matrix $C = [C^i_j]$ defined by

$$C^i_j = \left\langle A^i \mid B_j \right\rangle_{\mathbf{R}^p} = A^i_1 B^1_j + A^i_2 B^2_j + \cdots + A^i_p B^p_j = \sum_{k=1}^{p} A^i_k B^k_j.$$

That is, C^i_j is the dual pairing of the ith row of A, an element of $(\mathbf{R}^p)^*$, and the jth column of B, an element of $\mathbf{R}^p$.

Example 1.17. Let n be a positive integer, $(\boldsymbol{e}_i)_{i=1}^n$ the standard basis of $\mathbf{R}^n$, and $(\boldsymbol{e}^j)_{j=1}^n$ the standard dual basis of $(\mathbf{R}^n)^*$. For $1 \le i, j \le n$, we have $\boldsymbol{e}^j \boldsymbol{e}_i = \delta^j_i$, a 1×1 matrix. (In this "inner" product, $p = n$.)

Example 1.18. Let m and n be positive integers, $(\boldsymbol{e}_i)_{i=1}^m$ the standard basis of $\mathbf{R}^m$, and $(\boldsymbol{e}^j)_{j=1}^n$ the standard dual basis of $(\mathbf{R}^n)^*$. For each $i = 1, \dots, m$ and $j = 1, \dots, n$, the "outer" product $\boldsymbol{e}^j_i = \boldsymbol{e}_i \boldsymbol{e}^j$ (for which $p = 1$) is the $m \times n$ matrix having a 1 in the (i, j) entry and 0's everywhere else:

$$\boldsymbol{e}_i \boldsymbol{e}^j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 0 & \dots & 1 & \dots & 0 \end{bmatrix} = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix} = \boldsymbol{e}^j_i.$$

Matrix multiplication is associative, has an identity element, commutes with scalar multiplication, and distributes over addition. These properties allow us to work algebraically with matrices *as individual entities*, often avoiding the messiness of working with large collections of entries. To state these properties precisely requires only that we exercise care regarding sizes of operands.

Theorem 1.19. *Let m, p, q, n be positive integers, c a real number,*

$$A = [A^i_k] \in \mathbf{R}^{m\times p},\ B = [B^k_\ell],\ B' = [{B'}^k_\ell] \in \mathbf{R}^{p\times q},\ C = [C^\ell_j] \in \mathbf{R}^{q\times n}.$$

(i) $(AB)C = A(BC)$*;*

(ii) $I_m A = A$ *and* $AI_p = A$*;*

(iii) $A(cB) = c(AB)$*;*

(iv) $A(B + B') = AB + AB'$ *and* $(B + B')C = BC + B'C$*.*

Sketch of proof. In each part, it suffices to show that two matrices have the same (i, j) entries for all indices i and j.

(i). It suffices to prove $\sum_\ell (AB)^i_\ell C^\ell_j = \sum_k A^i_k (BC)^k_j$. But

$$\sum_{\ell=1}^{q} \left(\sum_{k=1}^{p} A^i_k B^k_\ell \right) C^\ell_j = \sum_{k=1}^{p} \sum_{\ell=1}^{q} A^i_k B^k_\ell C^\ell_j = \sum_{k=1}^{p} A^i_k \left(\sum_{\ell=1}^{q} B^k_\ell C^\ell_j \right).$$

(ii). For $1 \leq i \leq m$ and $1 \leq j \leq p$, we have

$$(AI_p)^i_j = \sum_{k=1}^{p} A^i_k \delta^k_j = A^i_j;$$

because of the factor δ^k_j, only the term with $k = j$ is non-zero. The other identity is similar.

(iii). $\displaystyle\sum_{k=1}^{p} A^i_k (cB^k_j) = \sum_{k=1}^{p} (cA^i_k) B^k_j.$

(iv). $\displaystyle\sum_{k=1}^{p} A^i_k (B^k_j + {B'}^k_j) = \sum_{k=1}^{p} (A^i_k B^k_j + A^i_k {B'}^k_j).$

The other distributive law is similar. □

Remark 1.20. In the proof of (ii), multiplying by a Kronecker delta and summing selects one term. This computational idiom is important.

Despite the "mixing" of entries in the definition of matrix multiplication, each identity comes down to a corresponding property for ordinary real number arithmetic.

Multiplication by a standard row or column "selects" part of A, yielding useful criteria for two matrices to be equal.

Corollary 1.21. *Let A be an $m \times n$ matrix, and let i and j be arbitrary indices with $1 \leq i \leq m$ and $1 \leq j \leq n$. Viewing $\boldsymbol{e}^i$ as a row matrix in $(\mathbf{R}^n)^*$ and $\boldsymbol{e}_j$ as a column matrix in $\mathbf{R}^m$:*

(i) *The product $\boldsymbol{e}^i A$ is A^i, the ith row of A;*

(ii) *The product $A\boldsymbol{e}_j$ is A_j, the jth column of A;*

(iii) *The product $\boldsymbol{e}^i A\boldsymbol{e}_j$ is A^i_j, the (i, j) entry of A.*

Particularly, if $A\boldsymbol{e}_j = \mathbf{0}^m$ for $j = 1, \dots, n$, then $A = \mathbf{0}^{m\times n}$.

Proof. Since $\boldsymbol{e}^i$ is the ith row of I_n and $\boldsymbol{e}_j$ is the jth column of I_m, (i), (ii), and (iii) follow immediately from Theorem 1.19 (ii). Finally, if $A\boldsymbol{e}_j$ is the zero vector for every j, then every entry of A is zero. □

Remark 1.22. To summarize, if $A\boldsymbol{x} = \mathbf{0}^m$ for all $\boldsymbol{x}$ in $\mathbf{R}^n$ (or even if this equation holds merely for the standard basis vectors), then A is the zero matrix. (The converse is obvious.)

Analogously, if $\boldsymbol{y}^\mathsf{T} A = (\mathbf{0}^n)^\mathsf{T}$ for all $\boldsymbol{y}$ in $\mathbf{R}^n$ (or even if $\boldsymbol{e}^i A = (\mathbf{0}^n)^\mathsf{T}$ for $i = 1, \dots, m$), then $A = \mathbf{0}^{m\times n}$.

Corollary 1.23. *If A and A' are $m \times n$ matrices, and if $A\boldsymbol{e}_j = A'\boldsymbol{e}_j$ for $j = 1, \dots, n$, then $A = A'$.*

Proof. Apply Corollary 1.21 to the difference $A - A'$, and use the distributive law to note that $(A - A')\boldsymbol{e}_j = A\boldsymbol{e}_j - A'\boldsymbol{e}_j$. □

Non-Commutativity of Matrix Multiplication

Matrix multiplication is not commutative. Even if A and B are both $n \times n$, so that both products AB and BA are defined and have the same size, it need not be the case that

$$(AB)^i_j = \sum_{k=1}^n A^i_k B^k_j \quad \text{and} \quad (BA)^i_j = \sum_{k=1}^n B^i_k A^k_j$$

are equal for all i and j.

Example 1.24. The matrices

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \text{ satisfy } AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Example 1.25. If A is an $n \times n$ matrix, then A commutes with matrices including I_n, A, A^2, and in general every "polynomial" in A.

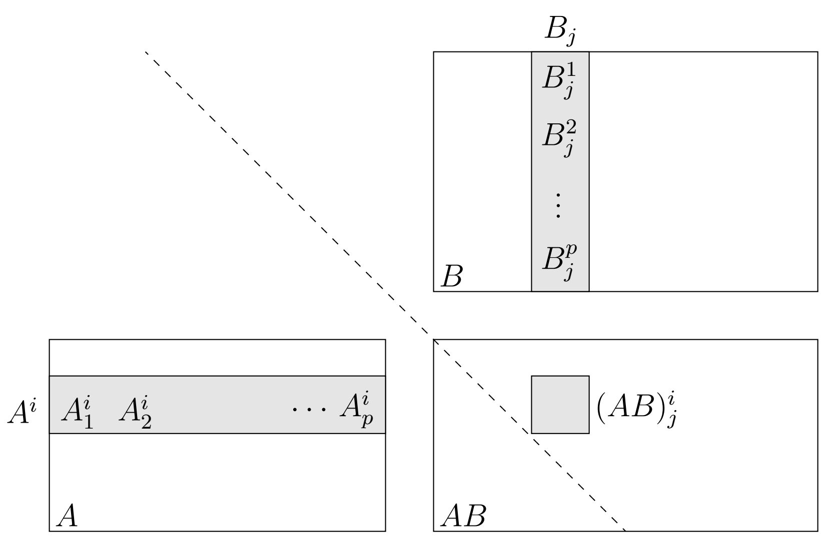
Theorem 1.26. *If $A = [A^i_k] \in \mathbf{R}^{m\times p}$ and $B = [B^k_j] \in \mathbf{R}^{p\times n}$, then $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$ as elements of $\mathbf{R}^{n\times m}$.*

Proof. Since $(AB)^i_j = \sum_k A^i_k B^k_j$, the (i, j) entry of $(AB)^\mathsf{T}$ is

$$(AB)^j_i = \sum_{k=1}^{p} A^j_k B^k_i = \sum_{k=1}^{p} B^k_i A^j_k = \sum_{k=1}^{p} (B^\mathsf{T})^i_k (A^\mathsf{T})^k_j,$$

the (i, j) entry of $B^\mathsf{T} A^\mathsf{T}$. □

Remark 1.27. A "visual" proof of Theorem 1.26 can be given by considering the following diagrammatic depiction of matrix multiplication:



Reflecting the paper across the diagonal line has the effect of transposing and reversing the order of the factors, and of transposing the product.

Invertibility of Square Matrices

Definition 1.28. Let n be a positive integer. An $n \times n$ matrix A is *invertible* if there exists an $n \times n$ matrix B such that $AB = I_n$ and $BA = I_n$. Such a matrix B is called an *inverse* of A.

Remark 1.29. Part (i) of the next result shows an invertible matrix has a unique inverse. We may therefore speak of *the* inverse of A, and denote it A^{-1}. Note that A^{-1} is invertible, and $(A^{-1})^{-1} = A$.

Theorem 1.30. *Let A, B, and B' be $n \times n$ matrices.*

(i) *If B and B' are inverses of A, then $B = B'$.*

(ii) *If $AB = I_n$, and if either A or B is invertible, then $BA = I_n$, so $B = A^{-1}$.*

(iii) *If A and B are invertible, then their product AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.*

(iv) *If A is invertible, then A^{T} is invertible, and $(A^{\mathsf{T}})^{-1} = (A^{-1})^{\mathsf{T}}$.*

Proof. Each assertion is a general consequence of associativity:

(i). By hypothesis, $BA = I_n$ and $AB' = I_n$. Consequently,

$$B = BI_n = B(AB') = (BA)B' = I_nB' = B'.$$

(ii). If $AB = I_n$ and A is invertible, multiplying on the left by A^{-1} and on the right by A gives

$$I_n = A^{-1}I_nA = A^{-1}(AB)A = (A^{-1}A)BA = BA.$$

If instead B is invertible, $I_n = BI_nB^{-1} = B(AB)B^{-1} = BA$.

(iii). By definition, an inverse of AB is any matrix C such that $(AB)C = I_n$ and $C(AB) = I_n$: It suffices to check $(AB)(B^{-1}A^{-1}) = I_n$ and $(B^{-1}A^{-1})(AB) = I_n$. By associativity of matrix multiplication,

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= \big((AB)B^{-1}\big)A^{-1} \\ &= \big(A(BB^{-1})\big)A^{-1} = (AI_n)A^{-1} = AA^{-1} = I_n.\end{aligned}$$

The other equality is similar.

(iv). By Theorem 1.26, $A^{\mathsf{T}}(A^{-1})^{\mathsf{T}} = (A^{-1}A)^{\mathsf{T}} = I_n^{\mathsf{T}} = I_n$ and $(A^{-1})^{\mathsf{T}}A^{\mathsf{T}} = (AA^{-1})^{\mathsf{T}} = I_n^{\mathsf{T}} = I_n$. □

1.2 Systems of Linear Equations

Definition 1.31. Let $A = [A^i_j]$ be a real $m \times n$ matrix. The equation $A\boldsymbol{x} = \mathbf{0}^m$, with $\boldsymbol{x} = (x^1, \dots, x^n)$ unknown, is called a (*homogeneous*) *linear system of m equations in n variables*.

If $\boldsymbol{b} = (b^1, \dots, b^m) \neq \mathbf{0}^m$ is a fixed element of $\mathbf{R}^m$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ is called a *non-homogeneous linear system of m equations in n unknowns*. If there exists an $\boldsymbol{x}_0$ such that $A\boldsymbol{x}_0 = \boldsymbol{b}$, the system is said to be *consistent*.

The matrix A is called the *coefficient matrix*. The $m \times (n+1)$ block matrix $[A \mid \boldsymbol{b}]$ is called the *augmented matrix*.

The zero vector $\boldsymbol{x} = \mathbf{0}^n$ always satisfies the homogeneous system. The general goal with a homogeneous system is to describe the set of solutions, and in particular, to determine whether the system has *non-trivial solutions* (i.e., solutions other than $\mathbf{0}^n$).

A non-homogeneous system may or may not be consistent. The solution set of a *consistent* non-homogeneous system is closely related to the soluton set of the associated homogeneous system, see Theorem 1.32.

An algorithm for solving linear systems is presented below; the output of the algorithm is an equivalent linear system (i.e., having precisely the same set of solutions) that can be solved (or shown to be inconsistent) by inspection.

Theorem 1.32. *Let A be an $m \times n$ real matrix and $\boldsymbol{b}$ an element of $\mathbf{R}^m$. If there exists a solution $\boldsymbol{x}_0$, i.e., an element of $\mathbf{R}^n$ satisfying $A\boldsymbol{x}_0 = \boldsymbol{b}$, then $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{x}_h$ for some $\boldsymbol{x}_h$ in $\mathbf{R}^n$ satisfying $A\boldsymbol{x}_h = \mathbf{0}^m$.*

Proof. By hypothesis, $A\boldsymbol{x}_0 = \boldsymbol{b}$. Let $\boldsymbol{x}_h$ be an arbitrary element of $\mathbf{R}^n$. If we write $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{x}_h$, then $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $A\boldsymbol{x} = A\boldsymbol{x}_0$, if and only if

$$\mathbf{0}^m = A\boldsymbol{x} - A\boldsymbol{x}_0 = A(\boldsymbol{x} - \boldsymbol{x}_0) = A\boldsymbol{x}_h. \qquad \square$$

Row Reduction

Fix an $m \times n$ matrix A and a column $\boldsymbol{b}$ in $\mathbf{R}^m$. (We allow $\boldsymbol{b} = \mathbf{0}^m$, and treat the homogeneous and non-homogeneous cases together.) Let A^i denote the ith row of A. The matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ is equivalent,

by equating components, to the system of m scalar equations

$$A^1\boldsymbol{x} = b^1, \qquad A^2\boldsymbol{x} = b^2, \quad \dots, \qquad A^m\boldsymbol{x} = b^m.$$

We begin by giving precise criteria for a coefficient matrix under which a system can be "solved by inspection".

Definition 1.33. Let A be a real $m \times n$ matrix with rows A^i.

A row A^i has a *leading* 1 if the row is not zero, and the first (leftmost) non-zero entry is 1.

The matrix A is in *row-echelon form* if:

(i) Every non-zero row of A has a leading 1;

(ii) Successive leading 1's occur further to the right. That is, if the leading 1's occur in positions $(1, k_1)$, $(2, k_2)$, …, (ℓ, k_ℓ), then the column indices increase: $k_1 < k_2 < \dots < k_\ell$;

(iii) Zero rows (if any) are collected at the bottom.

The matrix A is in *reduced row-echelon form* if A is in row-echelon form, and in addition each leading 1 is the only non-zero entry in its column.

Remark 1.34. Though an $m \times n$ matrix A has infinitely many row-echelon forms, the *reduced* row-echelon form of A is unique, see Theorem 1.42.

Remark 1.35. If a coefficient matrix is in row-echelon form, the number of leading 1's is at most $\min(m, n)$, the smaller of the number of variables and the number of equations.

Definition 1.36. If the leading 1's in the reduced row-echelon matrix of A have positions $(1, k_1)$, …, (ℓ, k_ℓ), the variables x_{k_1}, …, x_{k_ℓ} of the system $A\boldsymbol{x} = \boldsymbol{b}$ are said to be *basic*; the remaining variables are *free*.

Remark 1.37. A non-homogeneous system $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent (has no solution) if and only if the augmented matrix (reduced to row-echelon form) has a leading 1 in the last column; such a row corresponds to the inconsistent equation $0 = 1$.

When the augmented matrix of a consistent system is in reduced row-echelon form, the basic variables are expressed in terms of the free variables, and the free variables may take arbitrary real values.

A homogeneous system has a non-trivial solution if and only if there is at least one free variable. Particularly, if $m < n$ (more variables than equations), a homogeneous system $A\boldsymbol{x} = \boldsymbol{0}$ has a non-trivial solution.

Elementary Row Operations

Next we describe a set of transformations that do not change the solution set of a linear system, and suffice to bring the augmented matrix to reduced row-echelon form.

I. Add a multiple of one equation to another equation.

II. Multiply some equation by a non-zero real number.

III. Swap two equations.

Symbolically, if i and i' denote distinct indices, the preceding operations correspond to

I. Replacing $A^i\boldsymbol{x} = b^i$ with $(A^i + cA^{i'})\boldsymbol{x} = b^i + cb^{i'}$ for some real c.

II. Replacing $A^i\boldsymbol{x} = b^i$ with $cA^i\boldsymbol{x} = cb^i$ for some $c \neq 0$.

III. Exchanging $A^i\boldsymbol{x} = b^i$ and $A^{i'}\boldsymbol{x} = b^{i'}$.

Proposition 1.38. *Each of the preceding operations preserves the set of solutions.*

Proof. Operations II. and III. obviously do not change the set of solutions. To see the same is true for I., note that if $A^i\boldsymbol{x} = b^i$ and $A^{i'}\boldsymbol{x} = b^{i'}$, then

$$(A^i + cA^{i'})\boldsymbol{x} = b^i + cb^{i'} \quad \text{and} \quad A^{i'}\boldsymbol{x} = b^{i'};$$

that is, every solution of the system before I. is a solution of the system after I. Conversely, if

$$(A^i + cA^{i'})\boldsymbol{x} = b^i + cb^{i'} \quad \text{and} \quad A^{i'}\boldsymbol{x} = b^{i'};$$

then adding $-c$ times the second to the first gives

$$A^i\boldsymbol{x} = b^i \quad \text{and} \quad A^{i'}\boldsymbol{x} = b^{i'};$$

that is, every solution of the system after I. is a solution of the system before I. Conceptually, each of the three operations is reversible. □

A system of equations $A\boldsymbol{x} = \boldsymbol{b}$ can be recovered from its augmented matrix; there is no need to "carry along" the variables x when calculating. Performed on the augmented matrix, the preceding transformations are called *elementary row operations* of type I. (add a multiple of

one row to another row), II. (multiply a row by a non-zero number), or III. (swap two rows).

The row-reduction algorithm has two "stages", each recursive. The first stage puts the augmented matrix into row-echelon form; the second puts the augmented matrix in reduced row echelon form.

1.0. Find the leftmost column of coefficients that is not $\mathbf{0}^m$; call this the "active" column.

1.1. Pick a non-zero entry in the active column and use operations of type I. or II. to make that entry equal to 1. Then swap this row with the first row, so the active column has 1 as its topmost entry.

1.2. Use operations of type I. to make every entry in the active column (except for the topmost) equal to 0. Specifically, if the ith row of the active column contains α^i, add $-\alpha^i$ times the first row to the ith row.

1.3 Mentally cross out the first row. If there are rows remaining, start over at Step 0; otherwise the first stage is complete.

The algorithm terminates after finitely many steps because in each "round", the matrix being modified has one fewer rows and at least one fewer non-zero columns.

To put a row-echelon matrix into reduced row-echelon form, we need only make the entries "above" each leading 1 equal to 0.

2.1. Find the rightmost leading 1 and use row operations of type I. to cancel the non-zero entries "above" it (similarly to I.2).

2.2. Mentally cross out the last non-zero row. If there are more leading 1's, start over at Step 2.1; otherwise the algorithm is complete.

Example 1.39. Describe (recalling that superscripts are indices, not exponents) the solution set of the non-homogeneous linear system

$$\begin{aligned} 3x^1 + x^2 + 3x^3 &= -11, \\ 2x^1 \phantom{{}+x^2} + 4x^3 &= -10, \end{aligned} \qquad \left[\begin{array}{ccc|c} 3 & 1 & 3 & -11 \\ 2 & 0 & 4 & -10 \end{array}\right].$$

The augmented matrix at right is read off by inspection. The row reduction algorithm operates on this matrix.

Round 1 (Step 0) The first column is not $\mathbf{0}$, so becomes "active".

(Step 1) Our goal is to produce a 1 in the first column with elementary row operations. There are at least three strategies: Multiply the first row by 1/3; multiply the second row by 1/2; subtract the second row from the first.

As a practical matter, avoiding fractions usually means less work. Since the coefficients in the second row are all even, dividing the second row by 2 avoids fractions; do this, then swap the rows:

$$\left[\begin{array}{ccc|c} 3 & 1 & 3 & -11 \\ 2 & 0 & 4 & -10 \end{array}\right] \xrightarrow{\frac{1}{2}R^2} \left[\begin{array}{ccc|c} 3 & 1 & 3 & -11 \\ 1 & 0 & 2 & -5 \end{array}\right] \xrightarrow{R^1 \leftrightarrow R^2} \left[\begin{array}{ccc|c} 1 & 0 & 2 & -5 \\ 3 & 1 & 3 & -11 \end{array}\right].$$

(Step 2) Subtract multiples of the first row from the remaining row(s) to "clear" the active column:

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & -5 \\ 3 & 1 & 3 & -11 \end{array}\right] \xrightarrow{R^2 - 3R^1} \left[\begin{array}{ccc|c} 1 & 0 & 2 & -5 \\ 0 & 1 & -3 & 4 \end{array}\right].$$

(Step 3) Mentally cross out the first row, look at the remaining system of $(m-1) = 1$ equation(s) in $(n-1) = 2$ variables, and start over:

$$x^2 - 3x^3 = 4, \qquad \left[\begin{array}{ccc|c} 0 & 1 & -3 & 4 \end{array}\right].$$

Round 2 (Step 0) The (new) first non-zero column becomes active.

(Step 1) The "active" coefficient is already 1, and there are no additional rows, so the first stage is complete. In fact, the augmented matrix is already in *reduced* row echelon form, so nothing needs to be done for the second stage.

We have shown the original system is equivalent to (has the same set of solutions as)

$$\begin{aligned} x^1 \quad + 2x^3 &= -5, \\ x^2 - 3x^3 &= 4, \end{aligned} \qquad \left[\begin{array}{ccc|c} 1 & 0 & 2 & -5 \\ 0 & 1 & -3 & 4 \end{array}\right].$$

The basic variables are x^1 and x^2, while x^3 is free. To describe an arbitrary solution, write each basic variable in terms of the free variable(s):

$$\begin{aligned} x^1 &= -5 - 2x^3, \\ x^2 &= 4 + 3x^3, \end{aligned} \qquad \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} -5 - 2x^3 \\ 4 + 3x^3 \\ x^3 \end{bmatrix} = \begin{bmatrix} -5 \\ 4 \\ 0 \end{bmatrix} + x^3 \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}.$$

On the right, the general solution of the original system is in "parametric form". To each real value of the "parameter" x^3 there corresponds a solution $\boldsymbol{x}$ of the original system.

Remark 1.40. In the preceding example, $\boldsymbol{x}_0 = (-5, 4, 0)$ is a "particular" solution, and every scalar multiple of $(-2, 3, 1)$ is a homogeneous solution $\boldsymbol{x}_h$, as you should check, compare Theorem 1.32.

Example 1.41. Find the general solution of

$$\begin{aligned} 3x^1 + 12x^2 + x^3 + x^4 + 3x^5 &= 7, \\ 2x^1 + 8x^2 + x^3 + x^4 + x^5 &= 5, \\ x^1 + 4x^2 + 3x^3 + x^4 - x^5 &= 5, \\ -x^1 - 4x^2 + 3x^3 + - 2x^5 &= 1. \end{aligned}$$

The row reduction algorithm applied to the augmented matrix gives:

$$\left[\begin{array}{rrrrr|r} 3 & 12 & 1 & 1 & 3 & 7 \\ 2 & 8 & 1 & 1 & 1 & 5 \\ 1 & 4 & 3 & 1 & -1 & 5 \\ -1 & -4 & 3 & 0 & -2 & 1 \end{array}\right] \xrightarrow{R^1 \leftrightarrow R^3} \left[\begin{array}{rrrrr|r} 1 & 4 & 3 & 1 & -1 & 5 \\ 2 & 8 & 1 & 1 & 1 & 5 \\ 3 & 12 & 1 & 1 & 3 & 7 \\ -1 & -4 & 3 & 0 & -2 & 1 \end{array}\right]$$

$$\xrightarrow[R^4+R^1]{R^2-2R^1,\ R^3-3R^1,} \left[\begin{array}{rrrrr|r} 1 & 4 & 3 & 1 & -1 & 5 \\ 0 & 0 & -5 & -1 & 3 & -5 \\ 0 & 0 & -8 & -2 & 6 & -8 \\ 0 & 0 & 6 & 1 & -3 & 6 \end{array}\right] \xrightarrow[-\frac{1}{2}R^3]{R^2+R^4,} \left[\begin{array}{rrrrr|r} 1 & 4 & 3 & 1 & -1 & 5 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 4 & 1 & -3 & 4 \\ 0 & 0 & 6 & 1 & -3 & 6 \end{array}\right]$$

$$\xrightarrow[R^4-6R^2]{R^3-4R^2,} \left[\begin{array}{rrrrr|r} 1 & 4 & 3 & 1 & -1 & 5 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 1 & -3 & 0 \end{array}\right] \xrightarrow{R^4-R^3} \left[\begin{array}{rrrrr|r} 1 & 4 & 3 & 1 & -1 & 5 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad (*)$$

$$\xrightarrow{R^1-R^3} \left[\begin{array}{rrrrr|r} 1 & 4 & 3 & 0 & 2 & 5 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R^1-3R^2} \left[\begin{array}{rrrrr|r} 1 & 4 & 0 & 0 & 2 & 2 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

The starred matrix is in row-echelon form, with leading 1's in positions $(1, 1)$, $(2, 3)$, and $(3, 4)$. The basic variables are therefore x^1, x^3, and x^4; the remaining variables, x^2 and x^5, are free. To parametrize the solution space, write $x^1 = 2 - 4x^2 - 2x^5$, $x^3 = 1$, $x^4 = 3x^5$, so

$$\begin{bmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \\ x^5 \end{bmatrix} = \begin{bmatrix} 2 - 4x^2 - 2x^5 \\ x^2 \\ 1 \\ 3x^5 \\ x^5 \end{bmatrix} = \underbrace{\begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{x}_0} + \underbrace{x^2 \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x^5 \begin{bmatrix} -2 \\ 0 \\ 0 \\ 3 \\ 1 \end{bmatrix}}_{\boldsymbol{x}_h}.$$

Here again, $\boldsymbol{x}_0 = (2, 0, 1, 0, 0)$ is a particular solution, and the remaining portion is the general homogeneous solution.

Uniqueness of Reduced Row-Echelon Form

Theorem 1.42. *Let A be an $m \times n$ matrix. If A'_1 and A'_2 are matrices obtained by putting A into reduced row-echelon form, then $A'_1 = A'_2$.*

*Proof.** If an $m \times (k+1)$ matrix A is in reduced row-echelon form, then the $m \times k$ matrix B obtained by removing the $(k+1)$th column is in reduced row-echelon form: Removing the last column cannot create a leading 1 or remove a row of 0's.

We proceed by induction on the number of columns. If A has one column, the theorem is obvious.

Assume inductively for some $k \geq 1$ that every $m \times k$ matrix has a unique reduced row-echelon form. Let A be $m \times (k+1)$, and suppose A'_1 and A'_2 are reduced row-echelon forms of A. Let B, B'_1 and B'_2 be the $m \times k$ matrices obtained by removing the $(k+1)$th column of A, A'_1 and A'_2, respectively.

Performing a sequence of row operations and then removing the $(k+1)$th column has the same effect as removing the $(k+1)$th column and performing the same sequence of row operations. Consequently, B'_1 and B'_2 are reduced row-echelon forms of B, hence are equal by the inductive hypothesis. It suffices to prove A'_1 and A'_2 have the same $(k+1)$th column.

Because A'_1 and A'_2 are obtained from A by elementary row operations, the homogeneous systems $A'_1\boldsymbol{x} = \mathbf{0}$, and $A'_2\boldsymbol{x} = \mathbf{0}$ have the same solution sets. Let ℓ be the number of leading 1's in B'_1, i.e., the number of non-zero rows after deleting the $(k+1)$th column of A'_1.

If A'_1 has a leading 1 in the $(k+1)$th column, then the $(\ell+1)$th row of $A'_1\boldsymbol{x} = \mathbf{0}$ reads $x^{k+1} = 0$. Since $A'_2\boldsymbol{x} = \mathbf{0}$ has the same solution set, A'_2 also has a leading 1 in the $(k+1)$th column, necessarily in the $(\ell+1)$th row. Since each leading 1 is the only non-zero entry in its column, $A'_1 = A'_2$. (This is the only step where the hypothesis of *reduced* row-echelon form is used.)

If instead A'_1 does *not* have a leading 1 in the $(k+1)$th column, then there exists an $\boldsymbol{x} = (x^1, \dots, x^k, 1)$ satisfying $A'_1\boldsymbol{x} = \mathbf{0} = A'_2\boldsymbol{x}$, and therefore $(A'_1 - A'_2)\boldsymbol{x} = \mathbf{0}$. However, the first k columns of $A'_1 - A'_2$

*Adapted from Thomas Yuster, *The Reduced Row Echelon Form of a Matrix is Unique: A Simple Proof*, Mathematics Magazine, **57**, 2 (1984), 93–94.

are zero, so $\mathbf{0}^m = (A'_1 - A'_2)\boldsymbol{x}$ is the $(k+1)$th column of $A'_1 - A'_2$, and again $A'_1 = A'_2$.

In either case, we have shown that a general $m \times (k+1)$ matrix A has a unique reduced row-echelon form. This establishes the inductive step, and completes the proof. □

1.3 Criteria for Invertibility

The basic properties of invertibility "at the level of matrices" are given by Theorem 1.30. In practice, we need algorithmic criteria for detecting invertibility of specific matrices, and methods for calculating inverses of invertible matrices. These algorithms necessarily operate "at the level of entries".

A 1×1 matrix is essentially a real number, $A = [A_1^1]$, and is invertible if the single entry is non-zero, in which case $A^{-1} = [1/A_1^1]$. A general 2×2 matrix admits a similar criterion for invertibility and a formula for the inverse:

Proposition 1.43. *A 2×2 matrix $A = [A_j^i]$ is invertible if and only if the quantity $\Delta = A_1^1 A_2^2 - A_1^2 A_2^1$ is non-zero, in which case*

$$A = \begin{bmatrix} A_1^1 & A_2^1 \\ A_1^2 & A_2^2 \end{bmatrix}, \qquad A^{-1} = \frac{1}{\Delta}\begin{bmatrix} A_2^2 & -A_2^1 \\ -A_1^2 & A_1^1 \end{bmatrix}.$$

Proof. The rows are proportional if and only if $\Delta = 0$, if and only if the row-echelon form has a row of zeros, if and only if A is not invertible.

The formula is proven by verifying $AA^{-1} = I_2$ and $A^{-1}A = I_2$. □

Remark 1.44. This formula for the inverse of a 2×2 matrix should be memorized. It may help to write

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

The quantity Δ, the *determinant* of A, is studied in Chapter 3.

There are analogous formulas for larger matrices, but they are prohibitively complicated for practical use. Instead we develop a row-reduction algorithm for inverting an $n \times n$ matrix (or detecting that the matrix is not invertible).

Elementary Matrices

Definition 1.45. Let A be an $m \times n$ matrix. An $m \times m$ matrix E is an *elementary matrix* if the product EA gives the result of performing an elementary row operation on A.

We will show by construction that elementary matrices exist for each type of row operation. The following "principle" is useful for writing down elementary matrices in practice. The proof is immediate from the definition by taking $A = I_m$.

Proposition 1.46. *An elementary matrix is obtained by performing the corresponding row operation on the identity matrix I_m, and is invertible.*

Example 1.47. The following 3×3 matrices (i) add c times the third row to the second ($R^2 + cR^3$); (ii) multiply the second row by c (cR^2); and (iii) swap the second and third rows ($R^2 \leftrightarrow R^3$).

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

To see these matrices have the expected effect, multiply by a general matrix A, expressed as a block matrix whose entries are rows, e.g.,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A^1 \\ A^2 \\ A^3 \end{bmatrix} = \begin{bmatrix} A^1 \\ A^2 + cA^3 \\ A^3 \end{bmatrix}; \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} A^1 \\ A^2 \\ A^3 \end{bmatrix} = \begin{bmatrix} A^1 \\ A^3 \\ A^2 \end{bmatrix}.$$

To construct elementary matrices for each type of row operation, let $(\boldsymbol{e}_k)_{k=1}^m$ and $(\boldsymbol{e}^\ell)_{\ell=1}^m$ be the standard basis of $\mathbf{R}^m$ and the standard dual basis of $(\mathbf{R}^m)^*$.

If i and j are indices between 1 and m, recall that $\boldsymbol{e}_i^j = \boldsymbol{e}_i \boldsymbol{e}^j$ is the $m \times m$ matrix with a 1 in the (i, j) entry and 0's elsewhere.

Remark 1.48. The product $\boldsymbol{e}^j A$ is the jth row of A, and $\boldsymbol{e}_i^j A = \boldsymbol{e}_i \boldsymbol{e}^j A$ is the $m \times n$ matrix whose ith row is the jth row of A, and whose other entries are zero.

I. If c is real, then $E = I_m + c\boldsymbol{e}_i^j$ implements the type I. elementary row operation $R^i + cR^j$: If A is an arbitrary $m \times n$ matrix, then $(I_m + c\boldsymbol{e}_i^j)A = A + c\boldsymbol{e}_i^j A$ is the result of adding c times the jth row of A to the ith row of A.

The diagonal entries E_k^k are all 1, the (i,j) entry is c, and all other entries of E are 0. The inverse operation is to subtract c times the ith row from the jth row: $E^{-1} = (I_m + c\boldsymbol{e}_i^j)^{-1} = I_m - c\boldsymbol{e}_i^j$. Since $(\boldsymbol{e}_j^i)(\boldsymbol{e}_j^i) = \mathbf{0}^{m\times m}$, the inverse relationship

$$(I_m + c\boldsymbol{e}_i^j)(I_m - c\boldsymbol{e}_i^j) = I_m$$

is a difference of squares factorization, compare Exercise 1.4.

II. If $c \neq 0$ is real, then $E = I_m + (c-1)e_i^i$ implements the type II. elementary row operation cR^i: $(I_m + (c-1)e_i^i)A$ is the result of multiplying the ith row of A by c.

The matrix E is diagonal, with $E_i^i = c$ and all other diagonal entries E_k^k equal to 1. The inverse operation $\frac{1}{c}R^i$ is to multiply the ith row by $1/c$.

III. If $i \neq j$, then $E = I_m - (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}^i - \boldsymbol{e}^j)$ implements the Type III. elementary row operation $R^i \leftrightarrow R^j$: EA is the result of swapping the ith and jth rows of A.

If $\tau = (i\ j)$ is the transposition exchanging i and j, then $E = [\delta_k^{\tau(k)}]$: the (k,k) entry is 1 if $k \neq i$ and $k \neq j$, while the (i,j) and (j,i) entries are 1, and all other entries are 0. This matrix is its own inverse.

An Algorithm for Matrix Inversion

A sequence of row operations reducing a matrix A to echelon form A' may be viewed as a *factorization* $A' = E_r E_{r-1} \dots E_1 A$, or after rearranging, $A = E_1^{-1} \dots E_{r-1}^{-1} E_r^{-1} A'$, of A into a product of elementary matrices and a single row-echelon matrix. This observation leads to a criterion for invertibility, and to an algorithm for computing A^{-1}.

Theorem 1.49. *If A is an $m \times m$ matrix, the following are equivalent:*

(i) *A can be row-reduced to the identity matrix I_m.*

(ii) *A is invertible.*

Proof. If A can be row-reduced to the identity matrix, then A is a product of elementary matrices, and hence invertible.

Inversely, if A cannot be reduced to the identity matrix, then the reduced row-echelon form of A has fewer than m leading 1's, and therefore has a row of 0's. It follows that the homogeneous system $A\boldsymbol{x} = \mathbf{0}^m$ has a non-trivial solution $\boldsymbol{x}_0$. If B is an arbitrary $m \times m$ matrix, then $(BA)\boldsymbol{x}_0 = B(A\boldsymbol{x}_0) = B\mathbf{0}^m = \mathbf{0}^m \neq \boldsymbol{x}_0$, so $BA \neq I_m$. That is, A has no inverse. $\square$

The algorithm for computing A^{-1} uses this result. First form the $m \times (2m)$ matrix $[A \mid I_m]$ by "augmenting" with the identity matrix. Now row reduce the left-hand block, applying the row operations all the way across. If the left-hand block is row-reduced to I_m, the right-hand block is A^{-1}; if the left-hand block acquires a row of 0's, then A is not invertible.

This works because a sequence of row operations reducing A to the identity may be viewed as a factorization

$$E_r E_{r-1} \dots E_1 \left[A \mid I_m\right] = \left[I_m \mid B\right].$$

Equality of the right-hand block says $B = E_r E_{r-1} \dots E_1$; equality of the left-hand block says $BA = I_m$. But B is a product of elementary matrices, and therefore invertible. By Theorem 1.30 (ii), $AB = I_m$, so $B = A^{-1}$.

Exercises

Exercise 1.1. There are five elementary row operations on 2×2 matrices. List them, and for each, write down the corresponding elementary matrix E and its inverse E^{-1}, and multiply matrices to check that $EE^{-1} = I_2$. (Four of the "E"s will contain an arbitrary constant c.)

Exercise 1.2. Write down the 4×4 elementary matrices for the indicated operations.

(a) Adding c times the first or third row to the second row, or to the fourth row. (Four matrices in all.)

(b) Exchanging the first, second, or third row with the fourth row.

Exercise 1.3. Use the row-reduction algorithm to derive the formula for the inverse of a 2×2 matrix. (For simplicity, you may use the notation of Remark 1.44, and assume $a \neq 0$.)

Exercise 1.4. If i, j, k, and ℓ are indices between 1 and m, show that

$$\boldsymbol{e}_i^j \boldsymbol{e}_k^\ell = \delta_k^j \, \boldsymbol{e}_i^\ell = \begin{cases} \boldsymbol{e}_i^\ell & \text{if } j = k, \\ \mathbf{0}^{m\times m} & \text{if } j \neq k. \end{cases}$$

Referring to Example 1.24, use this formula to compute AB and BA, noting that $A = \boldsymbol{e}_2^1$ and $B = \boldsymbol{e}_1^2$.
Suggestion: Use the results of Examples 1.17 and 1.18.

Exercise 1.5. Suppose $A = [A_j^i]$ is an $m \times n$ matrix, $\boldsymbol{x} = [x^j]$ is an ordered n-tuple of "inputs", and $\boldsymbol{y} = A\boldsymbol{x} = [y^i]$ is the corresponding ordered m-tuple of "outputs" upon multiplication by A.

Show that the entry A_j^i is the "coefficient of sensitivity of y^i with respect to x^j", in the sense that if the jth input x^j is incremented by an amount Δx^j (and the remaining inputs are held fixed), then the ith output y^i changes by $\Delta y^i = A_j^i \, \Delta x^j$.

Exercise 1.6. Let A be $m \times p$ and B be $p \times n$.

(a) If the ith row of A is $\mathbf{0}$, what (if anything) can be deduced about the product AB? What if the jth column of A is $\mathbf{0}$?

(b) If the ith row of B is $\mathbf{0}$, what (if anything) can be deduced about the product AB? What if the jth column of B is $\mathbf{0}$?

Exercise 1.7. If $A = \sum_j a^j \boldsymbol{e}_j^j$ and $B = \sum_j b^j \boldsymbol{e}_j^j$ are $n \times n$ *diagonal* matrices, prove $AB = BA$. (Part of the exercise is to cope with reindexing, but be sure you understand this result conceptually.)

Exercise 1.8. An $n \times n$ matrix A is *symmetric* if $A^{\mathsf{T}} = A$. Prove that if A and B are symmetric $n \times n$ matrices, then AB is symmetric if and only if $BA = AB$.
Hint: Use Theorem 1.26.

Exercise 1.9. If $A = [A_j^i]$ and $B = [B_j^i]$ are $n \times n$ matrices, their *commutator* is defined to be $[A, B] = AB - BA$. In particular, A and B commute if and only if $[A, B] = \mathbf{0}^{n\times n}$.

(a) Show that the (i, j) entry of $[A, B]$ is

$$[A, B]_j^i = \sum_{k=1}^n (A_k^i B_j^k - A_j^k B_k^i).$$

(b) Show that if A, B_1, and B_2 are $n \times n$ matrices and c is real, then

$$[A, cB_1 + B_2] = c[A, B_1] + [A, B_2].$$

Suggestion: Work at the level of matrices, not at the level of entries.

Exercise 1.10. Let $A = [A^i_j]$ be an $n \times n$ matrix, and $\boldsymbol{e}^\ell_k = \boldsymbol{e}_k \boldsymbol{e}^\ell$ the $n \times n$ matrix with a 1 in the (k, ℓ) entry and 0's elsewhere.

(a) Show that $[A, \boldsymbol{e}^\ell_k] = \mathbf{0}^{n\times n}$ if and only if $A^i_k \delta^\ell_j = A^\ell_j \delta^i_k$ for all i and j. Hint: The (i, j) entry of $\boldsymbol{e}^\ell_k$ is $\delta^i_k \delta^\ell_j$.

(b) Show that $AB = BA$ for every B in $\mathbf{R}^{n\times n}$ if and only if A is a scalar matrix, i.e., there exists a real number c such that $A = cI_n$. Hint: One direction is easy. For the converse, use part (a) and the fact that A commutes with *every* $\boldsymbol{e}^\ell_k$.

Exercise 1.11. Let $A = [A^i_j]$ be an $m \times n$ matrix, and let A_j denote the jth column.

(a) Write down three types of of "elementary column operation" analogous to elementary row operations.

(b) For each type of column operation, find an $n \times n$ matrix E such that AE is the result of performing that operation on A.

(c) Does there exist an $m \times m$ matrix E such that EA implements an elementary column operation? Explain.

Exercise 1.12. Let $n \geq 1$ be an integer. Show that the set $GL(n, \mathbf{R})$ of invertible, $n \times n$ real matrices is a group under matrix multiplication.* Suggestion Use Theorem 1.30.

Exercise 1.13. Let σ be a bijection of the set $\{1, 2, \dots, n\}$, i.e., a permutation. The matrix

$$\mathsf{P}_\sigma = \sum_{i=1}^{n} \boldsymbol{e}_i \boldsymbol{e}^{\sigma(i)} = \sum_{j=1}^{n} \boldsymbol{e}_{\sigma^{-1}(j)} \boldsymbol{e}^j,$$

whose (i, j) entry is

$$\delta^i_{\sigma(i)} = \delta^{\sigma^{-1}(j)}_j = \begin{cases} 1 & \text{if } j = \sigma(i), \\ 0 & \text{if } j \neq \sigma(i), \end{cases}$$

*The notation $GL(n, \mathbf{R})$ stands for the *general linear group* of size n, with real entries.

and which therefore has precisely one 1 in every row and column, is the associated *permutation matrix*.

(a) Write out the two 2×2 permutations and the associated matrices, and the six 3×3 permutations and associated matrices.

(b) If $\boldsymbol{x} = \sum_j x^j \boldsymbol{e}_j$, show that $\mathsf{P}_\sigma \boldsymbol{x} = \sum_j x^{\sigma(j)} \boldsymbol{e}_j$. Verify explicitly (by multiplying matrices) for the six 3×3 permutation matrices.

(c) Prove that P_σ is invertible, and that $(\mathsf{P}_\sigma)^{-1} = \mathsf{P}_{\sigma^{-1}} = \mathsf{P}_\sigma^{\mathsf{T}}$.

(d) Prove that if σ and τ are permutations, then $\mathsf{P}_\tau \mathsf{P}_\sigma = \mathsf{P}_{\tau\sigma}$. (That is, the mapping $\sigma \mapsto \mathsf{P}_\sigma$ is an injective group homomomorphism from the symmetric group to $GL(n, \mathbf{R})$.)

Exercise 1.14. Let $A = [A^i_j]$ be 2×2, $\boldsymbol{b} = (b^1, b^2)$, $\boldsymbol{x} = (x^1, x^2)$, and $\mathbf{0} = \mathbf{0}^{1\times 2}$. Evaluate the product

$$\begin{bmatrix} A^1_1 & A^1_2 & b^1 \\ A^2_1 & A^2_2 & b^2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \\ 1 \end{bmatrix} = \begin{bmatrix} A & \boldsymbol{b} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix}$$

in two ways: (i) as an ordinary product, and (ii) as a "product of block matrices". Show the results are compatible in an obvious sense.

Chapter 2

Vector Spaces

2.1 Vector Space Axioms

Definition 2.1. A *real vector space* $(V, +, \cdot)$ comprises a non-empty set V, a binary operation $+ : V \times V \to V$, and a *scalar multiplication* map $\cdot : \mathbf{R} \times V \to V$ satisfying, for all $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$ in V and all real numbers c_1, c_2,

(i) (Associativity of $+$) $(\boldsymbol{v}_1 + \boldsymbol{v}_2) + \boldsymbol{v}_3 = \boldsymbol{v}_1 + (\boldsymbol{v}_2 + \boldsymbol{v}_3)$;

(ii) (Commutativity of $+$) $\boldsymbol{v}_2 + \boldsymbol{v}_1 = \boldsymbol{v}_1 + \boldsymbol{v}_2$;

(iii) (Identity element) There exists a *zero vector* $\mathbf{0}$ in V such that

$$\mathbf{0} + \boldsymbol{v} = \boldsymbol{v} = \boldsymbol{v} + \mathbf{0} \qquad \text{for all } \boldsymbol{v} \text{ in } V;$$

(iv) (Additive inverses) For every $\boldsymbol{v}$ in V, there exists an element $-\boldsymbol{v}$ such that $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0} = (-\boldsymbol{v}) + \boldsymbol{v}$.

(v) ("Associativity" of $\cdot$) $(c_1c_2) \cdot \boldsymbol{v} = c_1 \cdot (c_2 \cdot \boldsymbol{v})$;

(vi) (Left distributivity) $(c_1 + c_2) \cdot \boldsymbol{v} = c_1 \cdot \boldsymbol{v} + c_2 \cdot \boldsymbol{v}$;

(vii) (Right distributivity) $c \cdot (\boldsymbol{v}_1 + \boldsymbol{v}_2) = c \cdot \boldsymbol{v}_1 + c \cdot \boldsymbol{v}_2$;

(viii) (Normalization) $1 \cdot \boldsymbol{v} = \boldsymbol{v}$.

Remark 2.2. The pair $(V, +)$ is an Abelian group, on which the real numbers "act" by multiplication. Generally, scalars may (for example) be complex numbers, and one might speak of a *complex* vector space. In this book, vector spaces are real unless the contrary is stated explicitly.

Remark 2.3. The operations $+$ and $\cdot$ are part of the definition of a vector space. However, it is often convenient (and not ambiguous) to speak of "a vector space V", suppressing explicit mention of the operations.

Remark 2.4. For emphasis, or when considering more than one vector space, we may write $\mathbf{0}^V$ to signify the zero vector of V.

Proposition 2.5. *Let* $(V, +, \cdot)$ *be a vector space. For every* $\boldsymbol{v}$ *in* V*:*

(i) $0 \cdot \boldsymbol{v} = \mathbf{0}$*;*

(ii) $(-1) \cdot \boldsymbol{v} = -\boldsymbol{v}$.

Proof. Since $0 + 0 = 0$, the left distributive law gives

$$0 \cdot \boldsymbol{v} + 0 \cdot \boldsymbol{v} = (0 + 0) \cdot \boldsymbol{v} = 0 \cdot \boldsymbol{v}.$$

Canceling $0 \cdot \boldsymbol{v}$ gives $0 \cdot \boldsymbol{v} = \mathbf{0}$.

Now, since $0 = 1 + (-1)$, the left distributive law gives

$$\mathbf{0} = 0 \cdot \boldsymbol{v} = \big(1 + (-1)\big) \cdot \boldsymbol{v} = 1 \cdot \boldsymbol{v} + (-1) \cdot \boldsymbol{v}.$$

But $1 \cdot \boldsymbol{v} = \boldsymbol{v}$ by normalization, so $\mathbf{0} = \boldsymbol{v} + (-1) \cdot \boldsymbol{v}$. By definition of additive inverses, $-\boldsymbol{v} = (-1) \cdot \boldsymbol{v}$. $\square$

Remark 2.6. The scalar multiplication dot is often omitted in practice. When calculating, we freely use $1\boldsymbol{v} = \boldsymbol{v}$, $0\boldsymbol{v} = \mathbf{0}$, and $(-1)\boldsymbol{v} = -\boldsymbol{v}$.

Function Spaces

Example 2.7. Let X be a non-empty set. The set $\mathcal{F}(X, \mathbf{R})$ of all real-valued functions on X becomes a vector space under "pointwise addition and scalar multiplication", i.e.,

$$(f + g)(x) = f(x) + g(x), \qquad (cf)(x) = c\big(f(x)\big).$$

The zero vector is the function whose value at each point is 0; the additive inverse of a function f is the function $(-f)(x) = -f(x)$. The axioms are immediate; a function f with domain X is an "X-tuple" of real numbers, i.e., a collection of real numbers $f(x)$ "indexed by" the points of X, and each vector space axiom corresponds to a property of addition or multiplication of real numbers.

For illustration, here is a proof of the right distributive law. If f_1 and f_2 are arbitrary functions on X and c is a real number, then the value of $c(f_1 + f_2)$ at x is

$$\begin{aligned}
\big[c(f_1 + f_2)\big](x) &= c\big[(f_1 + f_2)(x)\big] && \text{Defn. of scalar multiplication} \\
&= c\big[f_1(x) + f_2(x)\big] && \text{Defn. of vector addition} \\
&= cf_1(x) + cf_2(x) && \text{Distrib. law for real numbers} \\
&= (cf_1)(x) + (cf_2)(x) && \text{Defn. of scalar multiplication} \\
&= \big[(cf_1) + (cf_2)\big](x) && \text{Defn. of vector addition.}
\end{aligned}$$

Since the functions $c(f_1 + f_2)$ and $(cf_1) + (cf_2)$ have the same value at every x in X, they are the same function.

Example 2.8. If $\mathbf{N} = \{0, 1, 2, 3, \dots\}$ denotes the set of natural numbers, we write $\mathcal{F}(\mathbf{N}, \mathbf{R}) = \mathbf{R}^\omega$. An element of $\mathbf{R}^\omega$ is effectively an infinite ordered list of real numbers $(a_n)_{n=0}^{\infty}$, i.e., a *real sequence*.

Column Vectors and Matrices

Example 2.9. Let n be a positive integer, $X = \{1, 2, \dots, n\}$ a set with n elements. Because a function $f : X \to \mathbf{R}$ is essentially an ordered n-tuple of real numbers $f(1), \dots, f(n)$, we denote $\mathcal{F}(X, \mathbf{R})$ by $\mathbf{R}^n$.

We write elements of $\mathbf{R}^n$ as "column vectors". The vector space operations are defined *componentwise*:

$$\begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} + \begin{bmatrix} y^1 \\ \vdots \\ y^n \end{bmatrix} = \begin{bmatrix} x^1 + y^1 \\ \vdots \\ x^n + y^n \end{bmatrix}, \qquad c \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix} = \begin{bmatrix} cx^1 \\ \vdots \\ cx^n \end{bmatrix},$$

or simply $[x^j] + [y^j] = [x^j + y^j]$, $c[x^j] = [cx^j]$ for brevity. The zero vector of $\mathbf{R}^n$ is denoted $\mathbf{0}^n$.

Remark 2.10. The superscripts denote *row indices*, not exponents.

Remark 2.11. When we denote an element of $\mathbf{R}^n$ by $[x^j]$, "j" is a *dummy index*, having no meaning outside the brackets. We may use any convenient letter without changing the meaning: $[x^i] = [x^j] = [x^\alpha]$, etc.

When we speak of a component x^j, by contrast, j has a specific (possibly implicit) numerical value between 1 and n.

Example 2.12. If $X = \{1, 2, \dots, m\}$ and $Y = \{1, 2, \dots, n\}$, the Cartesian product $X \times Y$ consists of all ordered pairs (i, j) with $1 \leq i \leq m$ and $1 \leq j \leq n$. A real-valued function $f : X \times Y \to \mathbf{R}$ is essentially a doubly-indexed family of numbers $f(i, j) = A^i_j$, namely, a real $m \times n$ matrix. That is, $\mathcal{F}(X \times Y, \mathbf{R}) = \mathbf{R}^{m\times n}$.

Addition and scalar multiplication of functions correspond to the "entrywise" operations of matrix addition and scalar multiplication, so $(\mathbf{R}^{m\times n}, +, \cdot)$ is a vector space.

The *zero matrix* $\mathbf{0}^{m\times n}$ in $\mathbf{R}^{m\times n}$ is the identity element for addition. The *negative* of a matrix $A = [A^i_j]$ is $-A = [-A^i_j]$.

Polynomials

Example 2.13. Let n be a non-negative integer, t an indeterminate, and let t^n denote the nth power of t. If a_0, a_1, …, a_n are real numbers, the expression

$$p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$$

is called a *polynomial* with *coefficients* a_0, …, a_n. The *degree* of p is the largest index with non-zero coefficient. If all coefficients are 0, we define the degree to be $-\infty$.

We define *polynomial addition* and *scalar multiplication* by adding or multiplying coefficients. That is, if $q(t) = b_0 + b_1 t + \cdots + b_n t^n$, and if c is real, we define

$$\begin{aligned} (p + q)(t) &= (a_0 + b_0) + (a_0 + b_0)t + \cdots + (a_n + b_n)t^n, \\ (cp)(t) &= ca_0 + ca_1 t + ca_2 t^2 + \cdots + ca_n t^n. \end{aligned}$$

The set of all polynomials of degree at most n is denoted P_n. It is straightforward (if somewhat tedious) to check all the vector space axioms. When we study "subspaces", we will see that only two conditions require verification, and these are encoded by the fact that a sum or scalar multiple of polynomials is itself a polynomial.

Example 2.14. If a_0, a_1, …, a_n and b_1, …, b_n are real numbers, the expressions

$$\begin{aligned} C(t) &= a_0 + a_1 \cos t + a_2 \cos(2t) + \ldots + a_n \cos(nt), \\ S(t) &= \phantom{a_0 + {}} b_1 \sin t + b_2 \sin(2t) + \ldots + b_n \sin(nt), \end{aligned}$$

are called the *cosine polynomial* with *coefficients* $a_0, \ldots, a_n$ and the *sine polynomial* with coefficients $b_1, \ldots, b_n$.

Two cosine polynomials are added, and a cosine polynomial is multiplied by a scalar, as if they were functions on $X = \mathbf{R}$. The set of cosine polynomials turns out to be a vector space under these operations; the proof boils down to the fact that a sum of cosine polynomials is a cosine polynomial, and a scalar multiple of a cosine polynomial is a cosine polynomial.

Corresponding remarks for sine polynomials are true.

The Geometry of Vector Operations

The vector space $(\mathbf{R}^2, +, \cdot)$ may be viewed as the *Cartesian plane*, a plane equipped with a distinguished pair of perpendicular lines (the *coordinate axes*) meeting at the *origin*. A point of $\mathbf{R}^2$ is an ordered pair $\boldsymbol{x} = (x^1, x^2)$, and is located by treating x^1 as a horizontal (east-west) position and x^2 as a vertical (north-south) position. The origin is the zero vector, $\mathbf{0} = (0, 0)$.

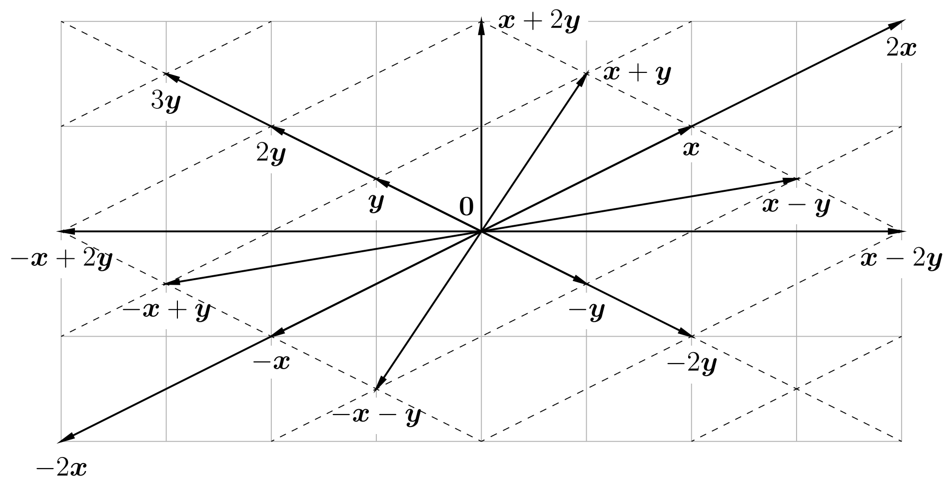


Figure 2.1: Vector addition and scalar multiplication in $\mathbf{R}^2$.

As a vector, $\boldsymbol{x}$ may be viewed as an arrow with tail at $\mathbf{0}$ and tip at $\boldsymbol{x}$. A vector sum $\boldsymbol{x}+\boldsymbol{y}$ is visualized by the *parallelogram law*: construct the parallelogram with a vertex at $\mathbf{0}$ and having sides $\boldsymbol{x}$ and $\boldsymbol{y}$; the fourth corner is $\boldsymbol{x}+\boldsymbol{y}$. Scalar multiplication $c\boldsymbol{x}$ is visualized by "scaling" $\boldsymbol{x}$

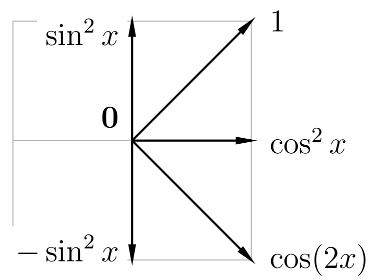
by a factor of c.

Example 2.15. In Figure 2.1, $\boldsymbol{x} = (2, 1)$ and $\boldsymbol{y} = (-1, \frac{1}{2})$ are plotted on a Cartesian grid of unit squares, along with several sums of scalar multiples. For example, $\boldsymbol{x} + \boldsymbol{y} = (1, \frac{3}{2})$ and $\boldsymbol{x} + 2\boldsymbol{y} = (2, 0)$.

Example 2.16. Analogous pictures can be drawn in more abstract vector spaces. For example, let $V = \mathcal{F}(\mathbf{R}, \mathbf{R})$ be the set of real-valued functions on $\mathbf{R}$, regarded as a vector space with ordinary addition and scalar multiplication of functions. The functions $f_1(x) = \cos^2 x$, $f_2(x) = \sin^2 x$, $g_1(x) = 1$, and $g_2(x) = \cos(2x)$ may each be viewed as an arrow, and any two of these functions "span a plane".

The trigonometric identities

$$\cos^2 x + \sin^2 x = 1,$$
$$\cos^2 x - \sin^2 x = \cos(2x)$$



may be interpreted schematically as vector sums. In particular, all four functions lie in a plane in V, since $g_1 = f_1 + f_2$ and $g_2 = f_1 - f_2$.

The identities $\cos^2 x = \frac{1}{2}(1 + \cos 2x)$ and $\sin^2 x = \frac{1}{2}(1 - \cos 2x)$ may also be interpreted using this diagram. (Check the details yourself.)

Linear Combinations

Definition 2.17. Let $(V, +, \cdot)$ be a vector space, and $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\dots$, $\boldsymbol{v}_p$ be distinct elements of V. If x^1, x^2, $\dots$, x^p are real numbers, the expression

$$x^1\boldsymbol{v}_1 + x^2\boldsymbol{v}_2 + \cdots + x^p\boldsymbol{v}_p = \sum_{j=1}^{p} x^j \boldsymbol{v}_j,$$

representing an element of V, is called the *linear combination* of the $\boldsymbol{v}_j$ with *coefficients* x^j. We say the linear combination is *non-trivial* if at least one coefficient x^j is non-zero.

Remark 2.18. If a coefficient is zero, the corresponding summand may as well be omitted. With this convention, a "non-trivial" linear combination may be regarded as a "non-empty" linear combination, i.e., as having at least one summand. An empty linear combination has value $\mathbf{0}^V$.

Example 2.19. If $\boldsymbol{x} = [x^j] \in \mathbf{R}^n$, and $(\boldsymbol{e}_j)_{j=1}^n$ denotes the standard basis of $\mathbf{R}^n$ (Definition 1.7), then

$$\boldsymbol{x} = \sum_{j=1}^n x^j \boldsymbol{e}_j, \quad \text{i.e.,} \quad \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix} = x^1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x^2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x^n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

In words, every column $\boldsymbol{x}$ in $\mathbf{R}^n$ is assembled as a linear combination from the standard basis using the components of $\boldsymbol{x}$ as coefficients.

Multiplying by a standard dual basis element $\boldsymbol{e}^k$ selects the kth component. By Theorem 1.19,

$$\boldsymbol{e}^k \boldsymbol{x} = \boldsymbol{e}^k \left(\sum_{j=1}^n x^j \boldsymbol{e}_j \right) = \sum_{j=1}^n x^j \boldsymbol{e}^k \boldsymbol{e}_j = \sum_{j=1}^n x^j \delta_j^k = x^k.$$

(Note the final step; summing $x^j \delta_j^k$ over j selects the term $j = k$.)

Example 2.20. Every row $\boldsymbol{\ell} = [\ell_i]$ in $(\mathbf{R}^n)^*$ may be written as a linear combination $\sum_i \ell_i \boldsymbol{e}^i$ of the standard dual basis $(\boldsymbol{e}^i)_{i=1}^n$.

Example 2.21. If $A = [A_1 \ \ \dots \ \ A_n]$ is an $m \times n$ matrix (partitioned into n columns, each an element of $\mathbf{R}^m$) and $\boldsymbol{x} = [x^j]$ is an element of $\mathbf{R}^n$, then the matrix product $A\boldsymbol{x} = \sum_j x^j A_j$, an element of $\mathbf{R}^m$, is precisely the linear combination of the columns of A with coefficients x^j.

Example 2.22. Recall that the product $\boldsymbol{e}_i^j = \boldsymbol{e}_i \boldsymbol{e}^j$ in $\mathbf{R}^{m \times n}$ has a 1 in the (i, j) entry and 0's elsewhere. If $A = [A_j^i]$ is an arbitrary element of $\mathbf{R}^{m \times n}$, then

$$A = \sum_{i=1}^m \sum_{j=1}^n A_j^i \boldsymbol{e}_i \boldsymbol{e}^j = \sum_{i=1}^m \sum_{j=1}^n A_j^i \boldsymbol{e}_i^j,$$

compare Example 2.19. In particular, $I_n = \displaystyle\sum_{i=1}^n \boldsymbol{e}_i \boldsymbol{e}^i = \sum_{i=1}^n \boldsymbol{e}_i^i$ in $\mathbf{R}^{n \times n}$.

Remark 2.23. A computational proof of the preceding claim illustrates a couple of algebraic idioms. Let A' denote the sum on the right. Since j is a dummy index in the expression for A', we may call it k.

Now, if j is an arbitrary index, $j = 1, \dots, n$, then

$$A' \boldsymbol{e}_j = \left(\sum_{i=1}^m \sum_{k=1}^n A_k^i \boldsymbol{e}_i \boldsymbol{e}^k \right) \boldsymbol{e}_j = \sum_{i=1}^m \sum_{k=1}^n A_k^i \boldsymbol{e}_i \delta_j^k = \sum_{i=1}^m A_j^i \boldsymbol{e}_i = A \boldsymbol{e}_j.$$

By Corollary 1.23, $A' = A$.

2.2 Subspaces

Definition 2.24. Let $(V, +, \cdot)$ be a vector space, and let $W \subseteq V$ be a non-empty subset.

If $\boldsymbol{x}+\boldsymbol{y} \in W$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in W, then W is *closed under addition*.

If $c\boldsymbol{x} \in W$ for all $\boldsymbol{x}$ in W and all real c, then W is *closed under scalar multiplication*.

If $(W, +, \cdot)$ is a vector space, we say W is a (*vector*) *subspace* of V.

Remark 2.25. If a non-empty set W in a vector space $(V, +, \cdot)$ is closed under scalar multiplication, then $\mathbf{0}^V \in W$: By hypothesis there is some $\boldsymbol{x}$ in W, so $0\boldsymbol{x} = \mathbf{0}^V \in W$. Further, W contains the additive inverse of each of its elements: If $\boldsymbol{x} \in W$, then $(-1) \cdot \boldsymbol{x} = -\boldsymbol{x} \in W$.

Theorem 2.26. *Let $(V, +, \cdot)$ be a vector space, $W \subseteq V$ non-empty. The following are equivalent:*

(i) *W is closed under addition and under scalar multiplication.*

(ii) *For all $\boldsymbol{x}$ and $\boldsymbol{y}$ in W and all real c, $c\boldsymbol{x} + \boldsymbol{y} \in W$.*

(iii) *W is a subspace of $(V, +, \cdot)$.*

Proof. ((i) if and only if (ii)). Assume W is closed under addition and scalar multiplication. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are elements of W and c is real, then $c\boldsymbol{x} \in W$ since W is closed under scalar multiplication, so $c\boldsymbol{x} + \boldsymbol{y} \in W$ since W is closed under addition.

Conversely, if Condition (ii) holds, then taking $c = 1$ shows W is closed under addition, while taking $\boldsymbol{y} = \mathbf{0}$ shows W is closed under scalar multiplication.

((i) if and only if (iii)). If W is closed under addition in V, then the addition operator $+$ is a binary operation on W itself, and is, *a fortiori*, associative and commutative on W.

If W is closed under scalar multiplication, W contains the identity element of V as well as the additive inverse of each of its elements, by Remark 2.25. That is, the first four vector space axioms hold for $(W, +, \cdot)$. The last four axioms hold automatically.

Conversely, if W is a subspace, then $(W, +, \cdot)$ is a vector space, so by definition W is closed under addition and scalar multiplication. $\square$

Remark 2.27. If $(V, +, \cdot)$ is a vector space, $W \subseteq V$ is a subspace, and if $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ are elements of W, then an arbitrary linear combination of the $\boldsymbol{v}_j$ is in W, by induction on the number of summands.

Example 2.28. Every vector space has two subspaces: The entire space (which is "not proper"), and the *trivial subspace* $W = \{\mathbf{0}^V\}$. A subspace other than these is "proper and non-trivial".

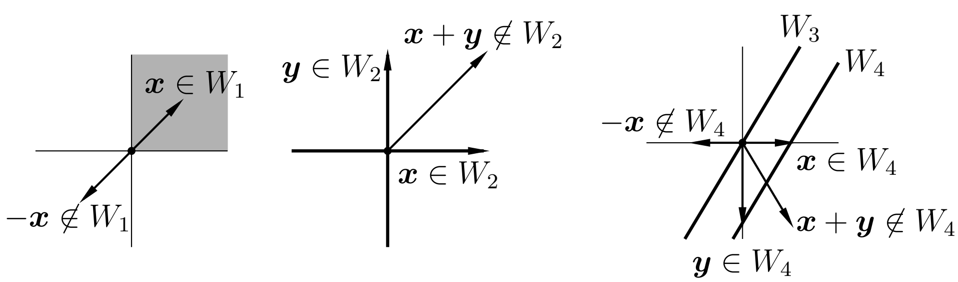
Example 2.29. Consider the vector space $(\mathbf{R}^2, +, \cdot)$, and write a general element $\boldsymbol{x}$ as (x^1, x^2).

The set $W_1 = \{\boldsymbol{x} : x^i \geq 0\}$, the *closed first quadrant*, is closed under addition, since if $\boldsymbol{x}$ and $\boldsymbol{y}$ are in W_1, then $\boldsymbol{x} + \boldsymbol{y} = (x^1 + y^1, x^2 + y^2)$ has non-negative components. (A sum of non-negative real numbers is non-negative.) This set is *not* closed under scalar multiplication. For example, $(1, 1) \in W_1$, but $-1(1, 1) = (-1, -1)$ is not in W_1.

The set $W_2 = \{\boldsymbol{x} : x^1x^2 = 0\}$, the union of the coordinate axes, is closed under scalar multiplication. If $\boldsymbol{x} \in W_2$ and c is real, then $c\boldsymbol{x} = (cx^1, cx^2)$ satisfies the membership criterion for W_2, since

$$(cx^1)(cx^2) = c^2(x^1x^2) = c^2(0) = 0.$$

This set is *not* closed under addition: $\boldsymbol{x} = (1, 0)$ and $\boldsymbol{y} = (0, 1)$ are in W_2, but their sum $\boldsymbol{x} + \boldsymbol{y} = (1, 1)$ is not in W_2.



The set $W_3 = \{\boldsymbol{x} : 5x^1 - 3x^2 = 0\}$ is closed under addition and scalar multiplication. To prove this, let $\boldsymbol{x}$ and $\boldsymbol{y}$ be arbitrary elements of W_3, and let c be real. We wish to show that $c\boldsymbol{x} + \boldsymbol{y}$ satisfies the defining equation of W_3. But

$$5(cx^1 + y^1) - 3(cx^2 + y^2) = c(5x^1 - 3x^2) + (5y^1 - 3y^2) = 0 + 0 = 0,$$

so $c\boldsymbol{x} + \boldsymbol{y} \in W_3$.

The set $W_4 = \{\boldsymbol{x} : 5x^1 - 3x^2 = 1\}$ is not closed under either addition or scalar multiplication. We give two proofs of each assertion. First, it suffices to find counterexamples. If we take $\boldsymbol{x} = (\frac{1}{5}, 0)$ and $\boldsymbol{y} = (0, -\frac{1}{3})$, then $\boldsymbol{x}$, $\boldsymbol{y} \in W_4$, but $\boldsymbol{x} + \boldsymbol{y} = (\frac{1}{5}, -\frac{1}{3})$ does not satisfy $5x^1 - 3x^2 = 1$, so $\boldsymbol{x} + \boldsymbol{y} \notin W_4$.

For scalar multiplication, the zero vector is not in W_4, so as remarked earlier, W_4 is not closed under scalar multiplication.

Alternatively, if $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfy the defining equation of W_4, we may add these equations, or multiply one of them by c:

$$\begin{aligned} 1 &= 5x^1 - 3x^2, & 1 &= 5x^1 - 3x^2, \\ 1 &= 5y^1 - 3y^2, & c &= c(5x^1 - 3x^2) \\ 2 &= 5(x^1 + y^1) - 3(x^2 + y^2), & &= 5(cx^1) - 3(cx^2). \end{aligned}$$

The left-hand side of the sum is $2 \neq 1$, so the pair $(x^1 + y^1, x^2 + y^2)$ is not in W_4. The left-hand side of the product is c, so if $c \neq 1$, the pair (cx^1, cx^2) is not in W_4.

Example 2.30. If a_1 and a_2 are real numbers (possibly 0), the set $W = \{\boldsymbol{x} : a_1x^1 + a_2x^2 = 0\} \subseteq \mathbf{R}^2$ is a subspace. Conceptually, the defining equation is "preserved by addition and by scalar multiplication".

Generally, if a_1, a_2, $\dots$, a_n are real numbers, the set

$$\{\boldsymbol{x} : a_1x^1 + a_2x^2 + \cdots + a_nx^n = 0\} \subseteq \mathbf{R}^n$$

is a subspace. This is easily verified directly, see Theorem 2.37.

Definition 2.31. Let n be a positive integer. A square matrix A in $\mathbf{R}^{n\times n}$ is said to be *symmetric* if $A^\mathsf{T} = A$, and is *skew-symmetric* if $A^\mathsf{T} = -A$.

Example 2.32. Let $V = \mathbf{R}^{n\times n}$ be the set of square matrices, regarded as a vector space under matrix addition and scalar multiplication. The set Sym^n of symmetric matrices in V is a subspace: If A and B are symmetric, and if c is real, then

$$(cA + B)^\mathsf{T} = cA^\mathsf{T} + B^\mathsf{T} = cA + B$$

by Remark 1.15. Theorem 2.26 implies Sym^n is a subspace of V. Similarly, the set Skew^n of skew-symmetric matrices in V is a subspace.

Example 2.33. Let $V = \mathbf{R}^{2\times 2}$ under addition and scalar multiplication.

The set $\mathbf{C}$ of matrices $[A^i_j]$ satisfying $A^1_1 = A^2_2$ and $A^2_1 = -A^1_2$ has typical element

$$\begin{bmatrix} a^1 & -a^2 \\ a^2 & a^1 \end{bmatrix} = a^1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + a^2 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = a^1 I + a^2 J,$$

where the final equality is the definition of J. If $A = a^1 I + a^2 J$ and $B = b^1 I + b^2 J$ are in $\mathbf{C}$, and if c is real, then

$$cA + B = c(a^1 I + a^2 J) + (b^1 I + b^2 J) = (ca^1 + b^1)I + (ca^2 + b^2)J$$

is in $\mathbf{C}$. By Theorem 2.26, $\mathbf{C}$ is a subspace of V.

The set W of matrices satisfying $A_1^1 A_2^2 - A_1^2 A_2^1 = 0$ is closed under scalar multiplication, but not under addition. For example, if

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad A + B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

then A and B are in W, but $A + B$ is not.

Example 2.34. Let $V = \mathcal{F}(\mathbf{R}, \mathbf{R})$, viewed as a vector space under ordinary addition and scalar multiplication of functions. The following subsets are vector subspaces, as is easily checked using Theorem 2.26.

$\mathcal{Z}(0) = \{f : f(0) = 0\}$, the set of functions vanishing at 0; generally, $\mathcal{Z}(x_0) = \{f : f(x_0) = 0\}$ if x_0 is real.

$\mathcal{C}(\mathbf{R}) = \{f : f \text{ is continuous}\}$, $\mathcal{D}(\mathbf{R}) = \{f : f \text{ is differentiable}\}$, $\mathcal{C}^1(\mathbf{R}) = \{f : f \text{ is continuously differentiable}\}$. In each case, a theorem from analysis shows that a sum or scalar multiple of functions of the stated type is another function of the same type.

P_n, the set of polynomial functions of degree at most n, and P, the space of all polynomial functions of arbitrary (finite) degree.

The space of all cosine polynomials, or of all sine polynomials.

$\{f \text{ in } \mathcal{C}^1 : f' = f\}$. This is an example of a *solution space* of a differential equation.

Theorem 2.35. *Let V be a vector space, W_1 and W_2 subspaces of V.*

(i) *The intersection $W_1 \cap W_2$ is a subspace of V.*

(ii) *The union $W_1 \cup W_2$ is a subspace of V if and only if $W_1 \subseteq W_2$ or $W_2 \subseteq W_1$.*

Proof. (i). Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be elements of $W_1 \cap W_2$, and let c be real. Since $\boldsymbol{x}$ and $\boldsymbol{y}$ are elements of W_1 and W_1 is a subspace, $c\boldsymbol{x} + \boldsymbol{y} \in W_1$. An entirely similar argument shows $c\boldsymbol{x} + \boldsymbol{y} \in W_2$. Thus $c\boldsymbol{x} + \boldsymbol{y} \in W_1 \cap W_2$. Since $\boldsymbol{x}$, $\boldsymbol{y}$, and c were arbitrary, $W_1 \cap W_2$ is a subspace.

(ii). If $W_1 \subseteq W_2$, then $W_1 \cup W_2 = W_2$ is a subspace. If $W_2 \subseteq W_1$, then $W_1 \cup W_2 = W_1$ is a subspace.

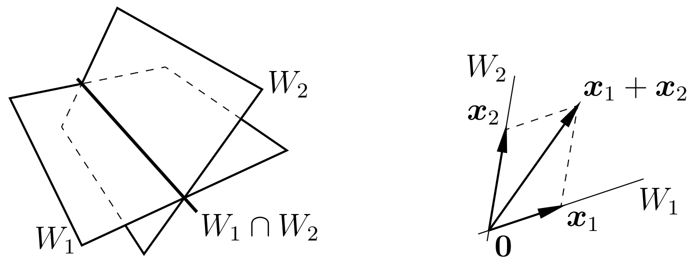


Figure 2.2: The intersection and union of subspaces.

Inversely, suppose neither subspace is contained in the other. Pick $\boldsymbol{x}_1$ in $W_1 \setminus W_2$, and pick $\boldsymbol{x}_2$ in $W_2 \setminus W_1$. Since each $\boldsymbol{x}_i$ is in $W_1 \cup W_2$, it suffices to show $\boldsymbol{x}_1 + \boldsymbol{x}_2$ is not in $W_1 \cup W_2$. But if $\boldsymbol{x}_1 + \boldsymbol{x}_2$ were an element of W_1, then $\boldsymbol{x}_2 = (\boldsymbol{x}_1 + \boldsymbol{x}_2) - \boldsymbol{x}_1$ would be an element of W_1, contrary to the choice of $\boldsymbol{x}_2$. Similarly, if $\boldsymbol{x}_1 + \boldsymbol{x}_2$ were an element of W_2, then $\boldsymbol{x}_1$ would be an element of W_2, contrary to choice. This means $\boldsymbol{x}_1 + \boldsymbol{x}_2$ is not in $W_1 \cup W_2$, so this set is not closed under addition. (The union *is* closed under scalar multiplication, as you can check.) □

Remark 2.36. An intersection of an arbitrary family of subspaces of V is a subspace of V, by an obvious modification of the proof of (i).

Theorem 2.37. *If $A = [A^i_j]$ is an $m \times n$ matrix, then*

$$W = \{[x^j] \text{ in } \mathbf{R}^n : A^i_1 x^1 + \cdots + A^i_n x^n = 0 \textit{ for all } i = 1, \dots, m\}$$

is a subspace of $(\mathbf{R}^n, +, \cdot)$.

We give two proofs.

Matrix multiplication. By the definition of matrix multiplication, the set W is the set of $\boldsymbol{x}$ in $\mathbf{R}^n$ such that $A\boldsymbol{x} = \mathbf{0}^m$. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are elements of W, and if c is real, then by properties of matrix multiplication (Theorem 1.19),

$$A(c\boldsymbol{x} + \boldsymbol{y}) = A(c\boldsymbol{x}) + A\boldsymbol{y} = c(A\boldsymbol{x}) + A\boldsymbol{y} = c(\mathbf{0}^m) + \mathbf{0}^m = \mathbf{0}^m.$$

That is, $c\boldsymbol{x} + \boldsymbol{y} \in W$, so W is a subspace by Theorem 2.26. □

Intersection of subspaces. For each $i = 1, \dots, m$, consider the set

$$W_i = \{[x^j] \text{ in } \mathbf{R}^n : A^i_1 x^1 + \cdots + A^i_n x^n = 0\}.$$

If $\boldsymbol{x} = [x^j]$ and $\boldsymbol{y} = [y^j]$ are in W_i and if c is real, then

$$\begin{aligned} &A^i_1(cx^1 + y^1) + \cdots + A^i_n(cx^n + y^n) \\ &\quad = c(A^i_1 x^1 + \cdots + A^i_n x^n) + (A^i_1 y^1 + \cdots + A^i_n y^n) = c \cdot 0 + 0 = 0. \end{aligned}$$

By Theorem 2.26, W_i is a subspace of $\mathbf{R}^n$ for $i = 1, 2, \dots, m$. By Theorem 2.35 (i), the intersection, namely W, is a subspace of $\mathbf{R}^n$. $\square$

Example 2.38. If $X \subseteq \mathbf{R}$ is an arbitrary non-empty subset, then

$$\mathcal{Z}(X) = \{f : \mathbf{R} \to \mathbf{R} : f(x) = 0 \text{ for all } x \text{ in } X\},$$

the set of functions vanishing identically on X, is a subspace of $\mathcal{F}(\mathbf{R}, \mathbf{R})$. This can be checked using Theorem 2.26. Alternatively, if x_0 is real, the set $\mathcal{Z}(x_0) = \{f : f(x_0) = 0\}$ is a subspace, so $\mathcal{Z}(X)$ is an intersection of subspaces:

$$\mathcal{Z}(X) = \bigcap_{x_0 \in X} \mathcal{Z}(x_0).$$

2.3 Spans, Sums of Subspaces

In the preceding section, subspaces of a vector space $(V, +, \cdot)$ are constructed by "cutting away", namely by imposing conditions on elements of V. Dually, a subspace can be "built up" by taking sums and scalar multiples of elements. This is the viewpoint explored below.

Definition 2.39. Let $(V, +, \cdot)$ be a vector space and $S \subseteq V$ a set of vectors. A *linear combination from* S is a finite sum of the form

$$x^1\boldsymbol{v}_1 + x^2\boldsymbol{v}_2 + \cdots + x^m\boldsymbol{v}_m = \sum_{i=1}^{m} x^i \boldsymbol{v}_i,$$

in which the vectors $\boldsymbol{v}_i$ in S are distinct, and the x^i are real numbers.

The *span* of S is the set $\operatorname{Span}(S)$ of all linear combinations from S. We say S *spans* V if $V = \operatorname{Span}(S)$. If some finite set S spans V, we say V is *finite-dimensional*.

Lemma 2.40. *Let $(V, +, \cdot)$ be a vector space, S and S' subsets of V.*

(i) $\operatorname{Span}(S)$ *is a subspace of* V.

(ii) *If* $S \subseteq S'$, *then* $\operatorname{Span}(S) \subseteq \operatorname{Span}(S')$.

(iii) $\operatorname{Span}\big(\operatorname{Span}(S)\big) = \operatorname{Span}(S)$.

Proof. (i). If S is empty, then $\operatorname{Span}(S) = \{\mathbf{0}^V\}$, which is a subspace. Otherwise, the fact that $\operatorname{Span}(S)$ is a subspace boils down to the fact that "a linear combination of linear combinations is a linear combination". Precisely, assume $\boldsymbol{x}$ and $\boldsymbol{y}$ are elements of $\operatorname{Span}(S)$, and let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ enumerate the vectors appearing in either linear combination. By hypothesis, there exist scalars x^i and y^i, $1 \leq i \leq m$, such that

$$\boldsymbol{x} = \sum_{i=1}^{m} x^i \boldsymbol{v}_i, \qquad \boldsymbol{y} = \sum_{i=1}^{m} y^i \boldsymbol{v}_i.$$

If c is real, then

$$c\boldsymbol{x} + \boldsymbol{y} = c\left(\sum_{i=1}^{m} x^i \boldsymbol{v}_i\right) + \sum_{i=1}^{m} y^i \boldsymbol{v}_i = \sum_{i=1}^{m} (cx^i + y^i)\boldsymbol{v}_i \in \operatorname{Span}(S).$$

By Theorem 2.26, $\operatorname{Span}(S)$ is a subspace.

(ii). This is essentially obvious: Every element of S is an element of S', so every linear combination from S is, *a fortiori*, a linear combination from S', i.e., $\operatorname{Span}(S) \subseteq \operatorname{Span}(S')$.

(iii). Clearly $S \subseteq \operatorname{Span}(S)$, so (ii) gives $\operatorname{Span}(S) \subseteq \operatorname{Span}\big(\operatorname{Span}(S)\big)$. Conversely, since $\operatorname{Span}(S)$ is a subspace, $\operatorname{Span}(S)$ is closed under addition and scalar multiplication. That is, an arbitrary linear combination from $\operatorname{Span}(S)$ is an element of $\operatorname{Span}(S)$, which proves the reverse inclusion, $\operatorname{Span}\big(\operatorname{Span}(s)\big) \subseteq \operatorname{Span}(S)$. □

Proposition 2.41. *Let $(V, +, \cdot)$ be a vector space, and $S \subseteq V$.* $\operatorname{Span}(S)$ *is the intersection of all subspaces of V that contain S.*

Proof. Let $\{W_\alpha\}$ denote the family of all subspaces of V that contain S, and let $W = \cap_\alpha W_\alpha$ be the intersection.

($W \subseteq \operatorname{Span}(S)$). Since $\operatorname{Span}(S)$ is itself a subspace containing S, i.e., is one of the W_α, we have $W \subseteq \operatorname{Span}(S)$.

($\operatorname{Span}(S) \subseteq W$). Let W_α be an arbitrary subspace of V such that $S \subseteq W_\alpha$. By Lemma 2.40, $\operatorname{Span}(S) \subseteq \operatorname{Span}(W_\alpha) = W_\alpha$. Since W_α is an arbitrary subspace containing S, $\operatorname{Span}(S) \subseteq \cap_\alpha W_\alpha = W$. □

Sums and Direct Sums of Subspaces

Definition 2.42. Let $(V, +, \cdot)$ be a vector space. If W_1 and W_2 are subspaces, their *sum* is

$$W_1 + W_2 = \{\boldsymbol{v} \text{ in } V : \boldsymbol{v} = \boldsymbol{w}_1 + \boldsymbol{w}_2 \text{ for some } \boldsymbol{w}_1 \text{ in } W_1 \text{ and } \boldsymbol{w}_2 \text{ in } W_2\}.$$

Proposition 2.43. *The sum $W_1 + W_2$ is a subspace of V.*

Proof. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be arbitrary elements of $W_1 + W_2$, and let c be real. By hypothesis, there exist elements $\boldsymbol{x}_1$ and $\boldsymbol{y}_1$ in W_1, and $\boldsymbol{x}_2$ and $\boldsymbol{y}_2$ in W_2, such that $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ and $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2$. Thus

$$c\boldsymbol{x} + \boldsymbol{y} = c(\boldsymbol{x}_1 + \boldsymbol{x}_2) + (\boldsymbol{y}_1 + \boldsymbol{y}_2) = (c\boldsymbol{x}_1 + \boldsymbol{y}_1) + (c\boldsymbol{x}_2 + \boldsymbol{y}_2).$$

Since W_1 is a subspace, $c\boldsymbol{x}_1 + \boldsymbol{y}_1 \in W_1$. Similarly, $c\boldsymbol{x}_2 + \boldsymbol{y}_2 \in W_2$. The preceding therefore shows $c\boldsymbol{x} + \boldsymbol{y} \in W_1 + W_2$. $\square$

Theorem 2.44. *Let V be a vector space. If S_1 and S_2 are subsets of V, and if $W_i = \operatorname{Span}(S_i)$, then $\operatorname{Span}(S_1 \cup S_2) = W_1 + W_2$.*

Proof. ($W_1 + W_2 \subseteq \operatorname{Span}(S_1 \cup S_2)$). Since $S_1 \subseteq S_1 \cup S_2$, Lemma 2.40 implies

$$W_1 = \operatorname{Span}(S_1) \subseteq \operatorname{Span}(S_1 \cup S_2).$$

Similarly, $W_2 \subseteq \operatorname{Span}(S_1 \cup S_2)$. Since $\operatorname{Span}(S_1 \cup S_2)$ is a subspace, we have $W_1 + W_2 \subseteq \operatorname{Span}(S_1 \cup S_2)$.

($\operatorname{Span}(S_1 \cup S_2) \subseteq W_1 + W_2$). If $\boldsymbol{x} \in \operatorname{Span}(S_1 \cup S_2)$, then there exist vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ in $S_1 \cup S_2$ and scalars x^i such that $\boldsymbol{x} = \sum_i x^i \boldsymbol{v}_i$. Reindexing if necessary, we may assume there is a k, $0 \leq k \leq m$, such that $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ are in S_1 and $\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_m$ are in S_2. Thus

$$\boldsymbol{x} = \left(\sum_{i=1}^{k} x^i \boldsymbol{v}_i\right) + \left(\sum_{i=k+1}^{m} x^i \boldsymbol{v}_i\right) \in W_1 + W_2. \qquad \square$$

Definition 2.45. Let $(V, +, \cdot)$ be a vector space, and let W_1 and W_2 be subspaces. If $W_1 \cap W_2 = \{\mathbf{0}^V\}$, we say $W_1 + W_2$ is a *direct sum*, and write $W_1 \oplus W_2$.

Theorem 2.46. *If $W = W_1 \oplus W_2$ is a direct sum of vector spaces, and if $\boldsymbol{w} \in W$, there exist* unique *vectors $\boldsymbol{w}_1$ in W_1 and $\boldsymbol{w}_2$ in W_2 such that $\boldsymbol{w} = \boldsymbol{w}_1 + \boldsymbol{w}_2$.*

Proof. By the definition of a sum of vector spaces, there exist vectors $\boldsymbol{w}_1$ in W_1 and $\boldsymbol{w}_2$ in W_2 satisfying $\boldsymbol{w} = \boldsymbol{w}_1 + \boldsymbol{w}_2$. To prove uniqueness, assume $\boldsymbol{w}_1'$ in W_1 and $\boldsymbol{w}_2'$ in W_2 provide another "decomposition" of $\boldsymbol{w}$. Since $\boldsymbol{w}_1 + \boldsymbol{w}_2 = \boldsymbol{w}_1' + \boldsymbol{w}_2'$, we have $\boldsymbol{w}_1 - \boldsymbol{w}_1' = \boldsymbol{w}_2' - \boldsymbol{w}_2$. However, the left-hand side is in W_1 and the right-hand side is in W_2. Since $W_1 \cap W_2 = \{\mathbf{0}^V\}$, we have $\boldsymbol{w}_1 - \boldsymbol{w}_1' = \mathbf{0}^V$ and $\boldsymbol{w}_2' - \boldsymbol{w}_2 = \mathbf{0}^V$, i.e., $\boldsymbol{w}_1 = \boldsymbol{w}_1'$ and $\boldsymbol{w}_2 = \boldsymbol{w}_2'$. □

Example 2.47. In $(\mathbf{R}^3, +, \cdot)$, the planes $W_1 = \{\boldsymbol{x} : x^1 = 0\}$ and $W_2 = \{\boldsymbol{x} : x^2 = 0\}$ are subspaces (Theorem 2.37) whose sum is all of $\mathbf{R}^3$. The general element $\boldsymbol{x} = (x^1, x^2, x^3)$ may be written

$$(0, x^2, x^3) + (x^1, 0, 0) = (0, x^2, 0) + (x^1, 0, x^3) \in W_1 + W_2.$$

The sum is not direct, since the non-zero vector $(0, 0, 1)$ is in $W_1 \cap W_2$.

Example 2.48. In $(\mathbf{R}^4, +, \cdot)$, the planes $W_1 = \{(x^1, x^2, 0, 0)\}$ and $W_2 = \{(0, 0, x^3, x^4)\}$ are subspaces whose sum is all of $\mathbf{R}^4$, and the sum is clearly direct.

Proposition 2.49. *Let n be a positive integer. If $\mathbf{R}^{n\times n}$ is the vector space of square matrices under matrix addition and scalar multiplication,* Sym^n *is the subspace of symmetric matrices, and* Skew^n *the subspace of skew-symmetric matrices, then*

$$\mathbf{R}^{n\times n} = \mathrm{Sym}^n \oplus \mathrm{Skew}^n.$$

In particular, every square matrix can be written uniquely as the sum of a symmetric matrix and a skew-symmetric matrix.

Proof. $(W_1 + W_2 = \mathbf{R}^{n\times n})$. Let A be an arbitrary $n \times n$ matrix. The matrices

$$A_{\mathrm{sym}} = \tfrac{1}{2}(A + A^{\mathsf{T}}), \qquad A_{\mathrm{skew}} = \tfrac{1}{2}(A - A^{\mathsf{T}})$$

are symmetric and skew-symmetric, respectively:

$$\begin{aligned} A_{\mathrm{sym}}^{\mathsf{T}} &= \tfrac{1}{2}(A + A^{\mathsf{T}})^{\mathsf{T}} = \tfrac{1}{2}\big(A^{\mathsf{T}} + (A^{\mathsf{T}})^{\mathsf{T}}\big) = \tfrac{1}{2}(A^{\mathsf{T}} + A) = A_{\mathrm{sym}}, \\ A_{\mathrm{skew}}^{\mathsf{T}} &= \tfrac{1}{2}(A - A^{\mathsf{T}})^{\mathsf{T}} = \tfrac{1}{2}\big(A^{\mathsf{T}} - (A^{\mathsf{T}})^{\mathsf{T}}\big) = \tfrac{1}{2}(A^{\mathsf{T}} - A) = -A_{\mathrm{skew}}. \end{aligned}$$

Moreover, $A_{\mathrm{sym}} + A_{\mathrm{skew}}$ is obviously A. This shows that an arbitrary $n \times n$ matrix can be written as the sum of a symmetric and a skew-symmetric matrix.

$(W_1 \cap W_2 = \{\mathbf{0}^{n\times n}\})$. If A is both symmetric and skew-symmetric, then $A = A^{\mathsf{T}} = -A$, so $A = \mathbf{0}^{n\times n}$. □

2.4 Bases and Dimension

The "size" of a real vector space is quantified by the minimum number of spanning elements. For example, $(\mathbf{R}^n, +, \cdot)$ is spanned by the standard basis $(\boldsymbol{e}_i)_{i=1}^n$. It turns out that every set of fewer than n vectors does not span, while every set of more than n vectors is "redundant", or "linearly dependent", in that some proper subset also spans. (It is not true that every set of n vectors spans $\mathbf{R}^n$; for example, one of the vectors could be $\mathbf{0}^n$, or all the vectors could be proportional.)

Linear Dependence and Independence

Definition 2.50. A set S of vectors in V is *linearly dependent* if there exists a non-trivial linear combination from S equal to the zero vector, i.e., if there exist distinct vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ in S ($k \geq 1$) and scalars x^i, *not all zero*, such that

$$\sum_{i=1}^{k} x^i \boldsymbol{v}_i = \mathbf{0}^V.$$

If S is not linearly dependent, we say S is *linearly independent*.

Remark 2.51. The empty set is linearly independent. A non-empty set S is linearly independent if and only if no non-trivial linear combination from S is $\mathbf{0}^V$. For later use we give a formal statement.

Lemma 2.52. *Let $(V, +, \cdot)$ be a vector space. A non-empty set S of vectors in V is linearly independent if and only if: For all $(\boldsymbol{v}_i)_{i=1}^k$ in S and all scalars $(x^i)_{i=1}^k$, if $\sum_i x^i \boldsymbol{v}_i = \mathbf{0}^V$, then $x^i = 0$ for all i.*

Example 2.53. In $(\mathbf{R}^2, +, \cdot)$, the set $S = \{(1, 0), (0, 1)\} = (\boldsymbol{e}_j)_{j=1}^2$ is linearly independent: If $(x^1, x^2) = x^1\boldsymbol{e}_1 + x^2\boldsymbol{e}_2 = (0, 0)$, then (by equating components) $x^1 = x^2 = 0$.

The set $S' = \{(1, 1), (1, -1), (4, 2)\} = (\boldsymbol{v}_j)_{j=1}^3$ is linearly dependent. For example, $3\boldsymbol{v}_1 + \boldsymbol{v}_2 - \boldsymbol{v}_3 = (0, 0)$.

Theorem 2.54. *Let $(V, +, \cdot)$ be a vector space, and let S be a set of vectors in V. The following are equivalent:*

(i) *S is linearly dependent.*

(ii) *There exists a vector $\boldsymbol{v}$ in S such that $\boldsymbol{v} \in \operatorname{Span}(S \setminus \{\boldsymbol{v}\})$. That is, some element of S can be expressed as a linear combination of the other elements of S.*

(iii) *There is a proper subset S' of S such that* $\mathrm{Span}(S') = \mathrm{Span}(S)$.

Proof. ((i) implies (ii)). If S is linearly dependent, there exist distinct vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ in S and scalars $x^1, \dots, x^k$, not all zero, such that

$$\mathbf{0} = x^1\boldsymbol{v}_1 + \cdots + \boldsymbol{x}^k\boldsymbol{v}_k.$$

Without loss of generality, we may assume $x^1 \neq 0$. The preceding equation may be rearranged as

$$\boldsymbol{v}_1 = -\frac{1}{x^1}(x^2\boldsymbol{v}_2 + \cdots + x^k\boldsymbol{v}_k),$$

which expresses some element of S as a linear combination of other elements.

((ii) implies (iii)). Let $\boldsymbol{v}$ be as in (ii), and let $S' = S \setminus \{\boldsymbol{v}\} \subseteq S$. By Lemma 2.40, $\mathrm{Span}(S') \subseteq \mathrm{Span}(S)$. Conversely, $\boldsymbol{v} \in \mathrm{Span}(S')$, so every linear combination from S is a linear combination from S', i.e., $\mathrm{Span}(S) \subseteq \mathrm{Span}(S')$.

((iii) implies (i)). Let S' be a proper subset of S with $\mathrm{Span}(S') = \mathrm{Span}(S)$, and let $\boldsymbol{v}_1$ be an element of $S \setminus S'$. Since $\boldsymbol{v}_1 \in \mathrm{Span}(S')$, there exist vectors $\boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ in S' (necessarily distinct from $\boldsymbol{v}_1$) and scalars $x^2, \dots, x^k$, such that

$$\boldsymbol{v}_1 = x^2\boldsymbol{v}_2 + \cdots + x^k\boldsymbol{v}_k, \quad \text{i.e.,} \quad -\boldsymbol{v}_1 + x^2\boldsymbol{v}_2 + \cdots + x^k\boldsymbol{v}_k = \mathbf{0}^V.$$

This is a non-trivial linear combination from S whose sum is $\mathbf{0}$. By definition, S is linearly dependent. □

Corollary 2.55. *Let S be a linearly independent subset of some vector space $(V, +, \cdot)$. If $\boldsymbol{v} \in V \setminus S$, then $S \cup \{\boldsymbol{v}\}$ is linearly independent if and only if $\boldsymbol{v} \notin \mathrm{Span}(S)$.*

Proof. If $\boldsymbol{v} \in \mathrm{Span}(S)$, then $\mathrm{Span}(S \cup \{\boldsymbol{v}\}) = \mathrm{Span}(S)$, so $S \cup \{\boldsymbol{v}\}$ is linearly dependent by the theorem. Contrapositively, if $S \cup \{\boldsymbol{v}\}$ is linearly independent, then $\boldsymbol{v} \notin \mathrm{Span}(S)$.

Suppose $\boldsymbol{v} \notin \mathrm{Span}(S)$, and assume

$$x\boldsymbol{v} + x^1\boldsymbol{v}_1 + \cdots + x^k\boldsymbol{v}_k = \mathbf{0}$$

for some vectors $\boldsymbol{v}_i$ in S and some scalars x, x^i. If $x \neq 0$, the preceding equation could be rearranged to express $\boldsymbol{v}$ as a linear combination from S, contrary to hypothesis; thus $x = 0$. Because S is linearly independent, $x^i = 0$ for $i = 1, \dots, k$. That is, the only linear combination from $S \cup \{\boldsymbol{v}\}$ equal to $\mathbf{0}$ is the trivial linear combination. By Lemma 2.52, $S \cup \{\boldsymbol{v}\}$ is linearly independent. □

Bases and Coordinate Vectors

Definition 2.56. An ordered, linearly independent spanning set of V is called a *basis* of V.

Remark 2.57. If $(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is a basis, then $(\boldsymbol{v}_2, \boldsymbol{v}_1)$ is a *different* basis. (Round brackets signify an ordered set, while curly braces signify an unordered set.) Many books do not explicitly specify that a basis is ordered, but ordering is implicit when basis elements are indexed.

Theorem 2.58. *Let $(V, +, \cdot)$ be a vector space spanned by some finite set T. If $S \subseteq V$ is a linearly independent set, there exists a basis S' with $S \subseteq S' \subseteq S \cup T$. In particular, some subset of T is a basis of V.*

Proof. The proof is inductive. Introduce the "initial" set $S_0 = S$, and write $T = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$. For $m = 1, \dots, n$, define

$$S_m = \begin{cases} S_{m-1} \cup \{\boldsymbol{y}_m\} & \text{if } \boldsymbol{y}_m \notin \operatorname{Span}(S_{m-1}), \\ S_{m-1} & \text{if } \boldsymbol{y}_m \in \operatorname{Span}(S_{m-1}). \end{cases}$$

By construction, $S_{m-1} \subseteq S_m \subseteq S \cup T$ and $\operatorname{Span}(\boldsymbol{y}_j)_{j=1}^m \subseteq \operatorname{Span}(S_m)$ for each m. In particular, $S \subseteq S_m \subseteq S \cup T$ for each m.

By Corollary 2.55 and induction on m, S_m is linearly independent.

The set $S' = S_n$ is linearly independent, satisfies $S \subseteq S' \subseteq S \cup T$, and spans V because $V = \operatorname{Span}(\boldsymbol{y}_j)_{j=1}^n \subseteq \operatorname{Span}(S') \subseteq V$. □

Theorem 2.59. *Let V be a vector space, and assume $S = (\boldsymbol{v}_i)_{i=1}^n$ is a basis of V. For every vector $\boldsymbol{x}$ in V, there exists a unique ordered n-tuple of scalars $(x^i)_{i=1}^n$ such that*

$$\boldsymbol{x} = \sum_{i=1}^n x^i \boldsymbol{v}_i.$$

Proof. (Existence). Let $\boldsymbol{x}$ be an arbitrary element of V. Existence of scalars x^i as in the theorem follows immediately from the definition of a spanning set.

(Uniqueness). Suppose there exist vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_m$ in S and scalars x^i and y^i such that $\boldsymbol{x} = \sum_i x^i \boldsymbol{v}_i$ and $\boldsymbol{x} = \sum_i y^i \boldsymbol{v}_i$. Subtracting the second from the first,

$$\mathbf{0}^V = \left(\sum_{i=1}^m x^i \boldsymbol{v}_i\right) - \left(\sum_{i=1}^m y^i \boldsymbol{v}_i\right) = \sum_{i=1}^m (x^i - y^i)\boldsymbol{v}_i.$$

By Lemma 2.52, $x^i - y^i = 0$ for all i, i.e., $x^i = y^i$ for all i. □

Definition 2.60. The ordered n-tuple $[\boldsymbol{x}]^S = [x^i]$ associated to the vector $\boldsymbol{x}$ with respect to the basis S is called the *coordinate vector* of $\boldsymbol{x}$ (in the basis S).

Example 2.61. Let $S = (\boldsymbol{e}_i)_{i=1}^n$ be the standard basis of $\mathbf{R}^n$,

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \qquad \dots, \qquad \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

If $\boldsymbol{x} = [x^i] \in \mathbf{R}^n$, then $\boldsymbol{x} = \sum_i x^i \boldsymbol{e}_i = [\boldsymbol{x}]^S$; every vector is its own coordinate vector in the standard basis.

Example 2.62. In $\mathbf{R}^{2\times 2}$, the matrices $(\boldsymbol{e}_i^j)_{i,j=1}^2$, namely,

$$\boldsymbol{e}_1^1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \boldsymbol{e}_1^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \boldsymbol{e}_2^1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \qquad \boldsymbol{e}_2^2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

constitute the *standard basis* of $\mathbf{R}^{2\times 2}$. An arbitrary matrix $A = [A_j^i]$ may be written uniquely as a linear combination

$$\begin{bmatrix} A_1^1 & A_2^1 \\ A_1^2 & A_2^2 \end{bmatrix} = A_1^1\boldsymbol{e}_1^1 + A_2^1\boldsymbol{e}_1^2 + A_1^2\boldsymbol{e}_2^1 + A_2^2\boldsymbol{e}_2^2 = \sum_{i,j=1}^{2} A_j^i \boldsymbol{e}_i^j.$$

Example 2.63. For each pair of indices (i, j) with $1 \le i \le m$ and $1 \le j \le n$, let $\boldsymbol{e}_i^j$ be the $m\times n$ matrix having $\boldsymbol{e}_i$ as its jth column and all other columns equal to $\mathbf{0}$, i.e., having a 1 in the ith row and jth column and 0 elsewhere. The collection $S = (\boldsymbol{e}_i^j)$ of all such matrices forms the *standard basis* of $\mathbf{R}^{m\times n}$. If $A = [A_j^i] \in \mathbf{R}^{m\times n}$, then

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} A_j^i \boldsymbol{e}_i^j.$$

That is, every $m \times n$ real matrix is its own coordinate vector with respect to the standard basis of $\mathbf{R}^{n\times m}$.

Dimension of a Vector Space

Theorem 2.64. *Let $(V, +, \cdot)$ be a vector space spanned by a set S containing m elements. If $S' \subseteq V$ is a linearly independent set of n elements, then $n \le m$.*

Proof. Write $S = (\boldsymbol{v}_i)_{i=1}^m$ and $S' = (\boldsymbol{v}'_j)_{j=1}^n$. Because S spans V, there exist scalars A^i_j, $1 \le i \le m$ and $1 \le j \le n$, such that

$$\boldsymbol{v}'_j = \sum_{i=1}^m A^i_j \boldsymbol{v}_i \quad \text{for all } j = 1, \dots, n.$$

These scalars constitute a matrix $A = [A^i_j]$ in $\mathbf{R}^{m\times n}$, which defines a homogeneous linear system $A\boldsymbol{x} = \mathbf{0}^m$ of m equations in n variables.

If $\boldsymbol{x} = [x^j]$ is an arbitrary vector of coefficients, then

$$\sum_{j=1}^n x^j \boldsymbol{v}'_j = \sum_{j=1}^n x^j \left(\sum_{i=1}^m A^i_j \boldsymbol{v}_i\right) = \sum_{i=1}^m \left(\sum_{j=1}^n A^i_j x^j\right) \boldsymbol{v}_i.$$

In particular, if $\boldsymbol{x}$ is a solution of $A\boldsymbol{x} = \mathbf{0}^m$, then the linear combination on the left is $\mathbf{0}^V$.

If $m < n$, i.e., the system $A\boldsymbol{x} = \mathbf{0}^m$ has more variables than equations, then there exists a non-trivial solution $\boldsymbol{x}$ by Remark 1.37. This means $\mathbf{0}^V$ is a non-trivial linear combination from S', i.e., S' is linearly dependent.

Contrapositively, if S' is linearly independent, then $n \le m$. □

Corollary 2.65. *Let $(V, +, \cdot)$ be a vector space admitting a basis consisting of m elements. If $S' = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ is a basis of V, then $n = m$.*

Proof. Let S be a basis of V containing m elements. Since S spans V and S' is linearly independent, the theorem gives $n \le m$. Since S' spans V and S is linearly independent, $m \le n$. □

Definition 2.66. If some (hence every) basis of V contains exactly n elements, we say V is *n-dimensional*, and write $n = \dim V$. If V has no finite basis, we say V is *infinite-dimensional*, and write $\dim V = \infty$.

Corollary 2.67. *If $(V, +, \cdot)$ is an n-dimensional vector space, and if W is a subspace, then $\dim W \le n$, with equality if and only if $W = V$.*

Proof. We first show W has a finite basis.

Let $S_0 = \varnothing$. If $W = \{\mathbf{0}^V\}$, then S_0 is a basis. Otherwise, define sets S_k (for $k = 1, 2, \dots$) inductively as follows: If $W \ne \operatorname{Span}(S_{k-1})$, pick a vector $\boldsymbol{v}_k$ in $W \setminus \operatorname{Span}(S_{k-1})$ and put $S_k = S_{k-1} \cup \{\boldsymbol{v}_k\}$. By construction, the set S_k is linearly independent and contains exactly k elements.

It must be that $W = \operatorname{Span}(S_m)$ for some integer m with $0 \le m \le n$; if not, then $S_{n+1} \subseteq V$ is a linearly independent set of $(n+1)$ elements, contrary to Theorem 2.64. This proves simultaneously that $\dim W$ is finite and no larger than n.

If $W = V$, then obviously $\dim W = \dim V$.

If $W \neq V$, there exists a vector $\boldsymbol{v}_0$ in $V \setminus W$; if $(\boldsymbol{v}_j)_{j=1}^m$ is a basis of W, then $(\boldsymbol{v}_j)_{j=0}^m$ is a linearly independent set in V by Corollary 2.55, so $\dim W = m < m+1 \le \dim V$. Contrapositively, if $\dim W = \dim V$, then $W = V$. □

Corollary 2.68. *If $(V, +, \cdot)$ is an n-dimensional vector space, and if S is a set of n elements, then S is linearly independent if and only if $\operatorname{Span}(S) = V$.*

Proof. Apply the preceding corollary to $W = \operatorname{Span}(S)$. □

Example 2.69. The standard basis of $(\mathbf{R}^n, +, \cdot)$ contains n elements, so $\dim \mathbf{R}^n = n$.

Example 2.70. The standard dual basis of $\big((\mathbf{R}^n)^*, +, \cdot\big)$ contains n elements, so $\dim(\mathbf{R}^n)^* = n$.

Example 2.71. The standard basis of $(\mathbf{R}^{m\times n}, +, \cdot)$ contains mn elements, so $\dim \mathbf{R}^{m\times n} = mn$.

Example 2.72. Let n be a non-negative integer. The polynomial space

$$P_n = \{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n\}$$

has the basis $(1, t, t^2, \dots, t^n)$, containing $(n+1)$ elements. (Superscripts denote exponents here.) Thus $\dim P_n = n + 1$.

Example 2.73. The vector space of symmetric 2×2 real matrices is 3-dimensional, and the space of skew-symmetric 2×2 real matrices is 1-dimensional. To show this, compare an arbitrary 2×2 matrix A to its transpose A^{T}:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad A^{\mathsf{T}} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

These are equal if and only if $b = c$. These are negatives if and only if $b = -c$ and $a = d = 0$. General symmetric and skew-symmetric 2×2 matrices can therefore be written

$$\begin{aligned} \begin{bmatrix} a & b \\ b & d \end{bmatrix} &= a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \qquad & \begin{bmatrix} 0 & -b \\ b & 0 \end{bmatrix} &= b \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\ &= a\boldsymbol{e}_1^1 + b(\boldsymbol{e}_1^2 + \boldsymbol{e}_2^1) + d\boldsymbol{e}_2^2, & &= b(\boldsymbol{e}_2^1 - \boldsymbol{e}_1^2). \end{aligned}$$

The right-hand sides are linear combinations of specific matrices, and clearly the only way the right-hand side can be the zero matrix is if every coefficient is 0. It follows that $\{\boldsymbol{e}_1^1, \boldsymbol{e}_1^2 + \boldsymbol{e}_2^1, \boldsymbol{e}_2^2\}$ is a basis for Sym^2, and $\{\boldsymbol{e}_2^1 - \boldsymbol{e}_1^2\}$ is a basis of Skew^2.

Example 2.74. The space P of all polynomial functions has infinite basis $\{1, t, t^2, t^3, \dots\}$, so $\dim P = \infty$.

Example 2.75. In the vector space $(\mathbf{R}^\omega, +, \cdot)$ of real sequences, define $\boldsymbol{e}_1 = (1, 0, 0, \dots)$, $\boldsymbol{e}_2 = (0, 1, 0, \dots)$, …, $\boldsymbol{e}_j = (0, \dots, 0, 1, 0, \dots)$, …, with $\boldsymbol{e}_j$ having a 1 in the jth component and 0s elsewhere. The set $S = (\boldsymbol{e}_j)_{j=1}^\infty$ is linearly independent, but *does not span* $\mathbf{R}^\omega$: Every linear combination from S is a *finite* sum, so every element of $\mathrm{Span}(S)$ has only finitely many non-zero components, while "most" elements of $\mathbf{R}^\omega$ have infinitely many non-zero components.

If we let $S^n = (\boldsymbol{e}_j)_{j=1}^n$ be the first n elements of S, we may identify $\mathbf{R}^n$ with $\mathrm{Span}(S^n)$, and with this identification we have inclusions $\mathbf{R}^1 \subseteq \mathbf{R}^2 \subseteq \ldots \subseteq \mathbf{R}^n \subseteq \ldots$. The union of these spaces, denoted $\mathbf{R}^\infty$, is $\mathrm{Span}(S)$. For every positive integer n, we have proper inclusions

$$\mathbf{R}^n \subsetneq \mathbf{R}^{n+1} \subsetneq \mathbf{R}^\infty \subsetneq \mathbf{R}^\omega.$$

Theorem 2.76. *The solution space of a homogeneous linear system $A\boldsymbol{x} = \mathbf{0}^m$ of m equations in n variables is a vector subspace of $\mathbf{R}^n$ whose dimension is equal to the number of free variables.*

Proof. The solution set of $A\boldsymbol{x} = \mathbf{0}^m$ is a subspace of $\mathbf{R}^n$ by Theorem 2.37. When the coefficient matrix is put into reduced row-echelon form, the resulting system $A'\boldsymbol{x} = \mathbf{0}^m$ has precisely the same solution space as the original system.

Let ℓ denote the number of free variables, and say the free variables are x^{k_1}, …, x^{k_ℓ}; for simplicity, write $x^{k_j} = y^j$. The basic variables are linear combinations of the free variables.

For each $j = 1, \dots, \ell$, let $\boldsymbol{v}_j$ be the solution obtained by setting $y^j = 1$ and all other free variables equal to 0. In particular, the k_j component of $\boldsymbol{v}_j$ is 1, and every other "free" solution has a 0 in the k_j component.

The solution space is clearly spanned by $(\boldsymbol{v}_j)_{j=1}^\ell$, and this set is linearly independent: If a^j are scalars satisfying $\mathbf{0}^n = a^1\boldsymbol{v}_1 + \cdots + a^\ell \boldsymbol{v}_\ell$, then $a^j = 0$, as the k_jth component of the linear combination. This shows the solution space has a basis of ℓ elements (ℓ the number of free variables). $\square$

Definition 2.77. If X is a vector space and $W \subseteq X$ is a subspace, a subspace X' is said to be *complementary* to W in X if $X = W \oplus X'$.

Lemma 2.78. *If X is a finite-dimensional vector space and W is a subspace, there exists a subspace X' complementary to W in X, and we have* $\dim X = \dim W + \dim X'$.

Proof. Write $k = \dim W$, $k + \ell = \dim X$. Pick a basis $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k)$ of W. By Theorem 2.58, there exist vectors $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell)$ such that $S = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_k, \boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell)$ is a basis of X.

Consider the subspace $X' = \operatorname{Span}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell)$. Since S spans X, $X = W + X'$. Since S is linearly independent, $W \cap X' = \{\mathbf{0}\}$. Finally, $\ell = \dim X'$, so $\dim X = k + \ell = \dim W + \dim X'$. □

Theorem 2.79 (The Dimension Theorem). *If X and Y are finite-dimensional subspaces of some vector space $(V, +, \cdot)$, then*

$$\dim(X + Y) = \dim X + \dim Y - \dim(X \cap Y).$$

Proof. Let $W = X \cap Y$, and pick a basis $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k)$ of W. Now use Lemma 2.78 to pick a subspace $X' \subseteq X$ with basis $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell)$ such that $X = W \oplus X'$, and a subspace $Y' \subseteq Y$, with basis $(\boldsymbol{y}_1, \dots, \boldsymbol{y}_m)$ such that $Y = W \oplus Y'$.

We claim that $S = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_k, \boldsymbol{x}_1, \dots, \boldsymbol{x}_\ell, \boldsymbol{y}_1, \dots, \boldsymbol{y}_m)$ is a basis of $X+Y$. Clearly $X+Y = \operatorname{Span}(S)$. To prove S is linearly independent, suppose there exist scalars $(a^i)_{i=1}^k$, $(b^i)_{i=1}^\ell$, and $(c^i)_{i=1}^m$ such that

$$\mathbf{0}^V = \sum_{i=1}^k a^i \boldsymbol{w}_i + \sum_{i=1}^\ell b^i \boldsymbol{x}_i + \sum_{i=1}^m c^i \boldsymbol{y}_i = \boldsymbol{w} + \boldsymbol{x} + \boldsymbol{y} \in W + X' + Y'.$$

Rearranging, $-\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{x}$; the left side is an element of $Y' \subseteq Y$ while the right side is in X. The common value is therefore an element of $W = X \cap Y$. But $W \cap Y' = \{\mathbf{0}\}$, so $\boldsymbol{y} = \mathbf{0}$ and $\boldsymbol{w} + \boldsymbol{x} = \mathbf{0}$. Since $W \cap X' = \{\mathbf{0}\}$, we have $\boldsymbol{w} = \mathbf{0}$ and $\boldsymbol{x} = \mathbf{0}$.

Each of the sets $(\boldsymbol{w}_i)_{i=1}^k$, $(\boldsymbol{x}_i)_{i=1}^\ell$, and $(\boldsymbol{y}_i)_{i=1}^m$ is linearly independent, and the summands $\boldsymbol{w}$, $\boldsymbol{x}$, and $\boldsymbol{y}$ are individually $\mathbf{0}$, so the scalars $(a^i)_{i=1}^k$, $(b^i)_{i=1}^\ell$, and $(c^i)_{i=1}^m$ are all zero. This shows that no non-trivial linear combination from S is $\mathbf{0}$. That is, S is linearly independent, hence is a basis for $X + Y$.

The dimension of $X + Y$ is found by counting basis elements:

$$\begin{aligned}\dim(X + Y) &= k + \ell + m = (k + \ell) + (k + m) - k \\ &= \dim X + \dim Y - \dim(X \cap Y).\end{aligned}$$ □

Example 2.80. A *plane* through the origin in $(\mathbf{R}^n, +, \cdot)$ is a two-dimensional subspace.

If W_1 and W_2 are distinct planes through the origin in $\mathbf{R}^3$, then $2 < \dim(W_1 + W_2) \leq \dim(\mathbf{R}^3) = 3$, so in fact $W_1 + W_2 = \mathbf{R}^3$ by Corollary 2.67. Theorem 2.79 implies $\dim(W_1 \cap W_2) = 1$. That is, distinct planes through the origin in $\mathbf{R}^3$ span $\mathbf{R}^3$, and intersect in a line through the origin.

A pair of distinct planes through the origin in $\mathbf{R}^4$ can intersect in a line (if the planes span a three-dimensional subspace) or a point (if the planes span $\mathbf{R}^4$).

Example 2.81. If W_1 and W_2 are distinct three-dimensional subspaces of $(\mathbf{R}^6, +, \cdot)$, their intersection can be a plane (if $\dim(W_1 + W_2) = 4$); a line (if $\dim(W_1 + W_2) = 5$); or a point (if $W_1 + W_2 = \mathbf{R}^6$).

Definition 2.82. An $n \times n$ matrix $[A^i_j]$ is:

(i) *Upper triangular* if every entry below the main diagonal is zero, i.e., if $A^i_j = 0$ whenever $i > j$.

(ii) *Lower triangular* if every entry above the main diagonal is zero, i.e., if $A^i_j = 0$ whenever $i < j$.

(i) *Diagonal* if every entry off the main diagonal is zero, i.e., if $A^i_j = 0$ whenever $i \neq j$.

Remark 2.83. The sets of upper triangular and lower triangular $n \times n$ matrices span $\mathbf{R}^{n \times n}$, and their intersection is the space of diagonal matrices.

Example 2.84. If $\boldsymbol{e}^j_i$ denotes the standard basis matrix with a 1 in the (i, j) entry and 0's elsewhere, then $\boldsymbol{e}^j_i$ is upper triangular if and only if $i \leq j$; lower triangular if and only if $j \leq i$; and diagonal if and only if $i = j$.

When $n = 2$, the space of diagonal matrices has basis $(\boldsymbol{e}^1_1, \boldsymbol{e}^2_2)$. In addition, $\boldsymbol{e}^2_1$ is upper triangular, and $\boldsymbol{e}^1_2$ is lower triangular. If we let D, U, and L denote the spaces of 2×2 diagonal, upper triangular, and lower triangular matrices, respectively, then $D = U \cap L$, $\mathbf{R}^{2\times 2} = U + L$, and

$$\dim \mathbf{R}^{2\times 2} = 4 = 3 + 3 - 2 = \dim U + \dim L - \dim D,$$

in agreement with Theorem 2.79.

Computational Algorithms

Let $(V, +, \cdot)$ be a vector space, and let $S = (\boldsymbol{v}_j)_{j=1}^n$ be an ordered subset. To determine whether S is linearly dependent or independent, express the zero vector $\mathbf{0}^V$ as a linear combination with coefficients $x^1, \dots, x^n$, reduce the resulting coefficient matrix to echelon form, and determine whether or not there are free variables (dependent) or not (independent).

Example 2.85. Is the set $S = \{(1, 1, -1, -1), (2, 2, 2, 2), (0, 0, 1, 1)\}$ linearly dependent? Here we might notice that $2\boldsymbol{v}_1 - \boldsymbol{v}_2 + 4\boldsymbol{v}_3 = \mathbf{0}^4$ by inspection, so S is linearly dependent. We can also proceed systematically. Suppose x^1, x^2, and x^3 are scalars such that $\sum_j x^j \boldsymbol{v}_j = \mathbf{0}$:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = x^1 \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} + x^2 \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + x^3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} x^1 + 2x^2 \\ x^1 + 2x^2 \\ -x^1 + 2x^2 + x^3 \\ -x^1 + 2x^2 + x^3 \end{bmatrix}$$

This is a homogeneous system, whose coefficient matrix is the 4×3 matrix whose columns are the elements of S. (That is, we need not have written out the system at all!) We have

$$\begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 0 \\ -1 & 2 & 1 \\ -1 & 2 & 1 \end{bmatrix} \xrightarrow[R_2 \leftrightarrow R_3]{\substack{R_2 - R_1, \\ R_4 - R_3,}} \begin{bmatrix} 1 & 2 & 0 \\ -1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{\substack{R_2 + R_1, \\ \frac{1}{4} R_2}} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

From the row-echelon form, we see that x^1 and x^2 are basic and x^3 is free. This system has a non-trivial solution (i.e., $\mathbf{0}^4$ is a non-trivial linear combination from S), so S is linearly dependent.

Example 2.86. For the S of the preceding example, is $\boldsymbol{b} = (1, 2, 3, 3)$ in $\mathrm{Span}(S)$? Here we wish to express $\boldsymbol{b}$ as a linear combination from S, i.e., to find scalars x^j such that $x^1\boldsymbol{v}_1 + x^2\boldsymbol{v}_2 + x^3\boldsymbol{v}_3 = \boldsymbol{b}$, or prove none exist. Letting A denote the coefficient matrix in the preceding example, we wish to solve $A\boldsymbol{x} = \boldsymbol{b}$. We form the augmented matrix and row reduce:

$$\left[\begin{array}{rrr|r} 1 & 2 & 0 & 1 \\ 1 & 2 & 0 & 2 \\ -1 & 2 & 1 & 3 \\ -1 & 2 & 1 & 3 \end{array}\right] \xrightarrow[R_2 \leftrightarrow R_3]{\substack{R_2 - R_1, \\ R_4 - R_3,}} \left[\begin{array}{rrr|r} 1 & 2 & 0 & 1 \\ -1 & 2 & 1 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Already we can stop; the third equation is inconsistent; the system $A\boldsymbol{x} = \boldsymbol{b}$ has no solutions, so $\boldsymbol{b} \notin \operatorname{Span}(S)$.

Example 2.87. In the (four-dimensional) polynomial space P_3, consider the set $S' = \{1 + t - t^2 - t^3, 2 + 2t + 2t^2 + 2t^3, t^2 + t^3\}$, and let $q(t) = 1 + 2t + 3t^2 + 3t^3$. (Superscripts on t denote exponents.) Is S' linearly independent? Is q an element of $\operatorname{Span}(S')$?

We use a^j to denote unknowns. (Superscripts on a connote indices.) We have two computational alternatives.

The first is to set $a^1 p_1 + a^2 p_2 + a^3 p_3 = 0$ as polynomials (i.e., to expand the left-hand side and equate all coefficients to 0), then solve the resulting system for the a^j, looking for a non-trivial solution; or to set $a^1 p_1 + a^2 p_2 + a^3 p_3 = q$ as polynomials and attempt to solve.

The second method is to pick any convenient basis of P_3, express each element of S' (and the polynomial q) as a coordinate vector (Definition 2.60), and to perform computations in $\mathbf{R}^4$. But $\{1, t, t^2, t^3\}$ is a basis of P_3, and the coordinate vectors of the elements of S' are precisely the elements of S in the preceding examples, while the coordinate vector of q is $\boldsymbol{b} = (1, 2, 3, 3)$. From our work in the two preceding examples, we learn that S' is linearly dependent, and $q \notin \operatorname{Span}(S')$.

Example 2.88. Let $S = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\} = (\boldsymbol{v}_j)_{j=1}^3$. Determine whether S is a basis of $\mathbf{R}^3$; if so, and if $\boldsymbol{b} = (b^1, b^2, b^3)$ is arbitrary, find the coordinate vector $(\boldsymbol{b})^S$.

Let x^1, x^2, and x^3 be coefficients. In theory our task is twofold: Show that if $\sum_j x^j \boldsymbol{v}_j = \mathbf{0}^3$, then $x^j = 0$ for all j (so that S is linearly independent, hence a basis of $\mathbf{R}^3$); and solve the system $\sum_j x^j \boldsymbol{v}_j = \boldsymbol{b}$. Both operations entail forming the coefficient matrix A whose columns are the elements of S and row reducing. Rather than row reduce the same matrix twice, we set out ("optimistically") to answer the second question, and along the way answer the first.

The augmented matrix is reduced as follows:

$$\left[\begin{array}{ccc|c} 0 & 1 & 1 & b^1 \\ 1 & 0 & 1 & b^2 \\ 1 & 1 & 0 & b^3 \end{array}\right] \xrightarrow[R_1 \leftrightarrow R_2]{\substack{R_3 - R_2, \\ R_3 - R_1,}} \left[\begin{array}{ccc|c} 1 & 0 & 1 & b^2 \\ 0 & 1 & 1 & b^1 \\ 0 & 0 & -2 & -b^1 - b^2 + b^3 \end{array}\right]$$

$$\xrightarrow[R_1 - R_3]{\substack{-\frac{1}{2}R_3, \\ R_2 - R_3,}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{1}{2}(b^2 + b^3 - b^1) \\ 0 & 1 & 0 & \frac{1}{2}(b^3 + b^1 - b^2) \\ 0 & 0 & 1 & \frac{1}{2}(b^1 + b^2 - b^3) \end{array}\right].$$

Because there are no free variables, the original set is linearly independent, hence is a basis. As a fringe benefit, we learn that

$$(\boldsymbol{b})^S = \frac{1}{2}\begin{bmatrix} b^2 + b^3 - b^1 \\ b^3 + b^1 - b^2 \\ b^1 + b^2 - b^3 \end{bmatrix} = \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix}.$$

That is $x^1\boldsymbol{v}_1 + x^2\boldsymbol{v}_2 + x^3\boldsymbol{v}_3 = \boldsymbol{b}$.

Example 2.89. Let $W \subseteq \mathbf{R}^5$ be spanned by $\boldsymbol{x}_1 = (1, 1, 1, 1, 1)$ and $\boldsymbol{x}_2 = (1, 2, 3, 4, 5)$. Find a homogeneous linear system $A\boldsymbol{x} = \mathbf{0}^5$ whose solution space is W.

The idea is to treat the *coefficients* of a scalar equation as unknowns; that is, suppose $A = [a^1 \quad a^2 \quad a^3 \quad a^4 \quad a^5]$ is a set of coefficients for a linear equation vanishing on W. The vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ satisfy $A\boldsymbol{x} = 0$ if and only if

$$\begin{aligned} a^1 + a^2 + a^3 + a^4 + a^5 &= 0, \\ a^1 + 2a^2 + 3a^3 + 4a^4 + 5a^5 &= 0. \end{aligned}$$

To solve, form the coefficient matrix and put it in reduced row-echelon form. The coefficient matrix has *rows* equal to the vectors in a basis of W:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \xrightarrow{R_2 - R_1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \xrightarrow{R_1 - R_2} \begin{bmatrix} 1 & 0 & -1 & -2 & -3 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}.$$

The basic variables are a^1 and a^2. Solving for the basic variables, we get $a^1 = a^3 + 2a^4 + 3a^5$, $a^2 = -2a^3 - 3a^4 - 4a^5$. The general solution is

$$\begin{bmatrix} a^1 \\ a^2 \\ a^3 \\ a^4 \\ a^5 \end{bmatrix} = \begin{bmatrix} a^3 + 2a^4 + 3a^5 \\ -2a^3 - 3a^4 - 4a^5 \\ a^3 \\ a^4 \\ a^5 \end{bmatrix} = a^3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + a^4 \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \\ 0 \end{bmatrix} + a^5 \begin{bmatrix} 3 \\ -4 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

The three columns on the right span the solution space; the corresponding linear equations vanish on W, and all together they "cut out" W. That is, W is the solution set of

$$\begin{aligned} x^1 - 2x^2 + x^3 \qquad\qquad &= 0, \\ 2x^1 - 3x^2 \qquad + x^4 \qquad &= 0, \\ 3x^1 - 4x^2 \qquad\qquad + x^5 &= 0; \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 2 & -3 & 0 & 1 & 0 \\ 3 & -4 & 0 & 0 & 1 \end{bmatrix}.$$

Exercises

Exercise 2.1. Let $V = \mathcal{F}(\mathbf{R}, \mathbf{R})$ be the vector space of real-valued functions on $\mathbf{R}$ under ordinary addition and scalar multiplication. Determine (with justification) which of the following sets are linearly independent.

(a) $\{1 - t + t^2, 1 + t + t^2, t + t^2\}$; $\{1 - t + t^2, 1 + t + t^2, 1 + t^2\}$.

(b) $\{\cos^2 t, \sin^2 t\}$; $\{1, \cos^2 t, \sin^2 t\}$.

(c) $\{\cos t, \sin t\}$; $\{1, \cos t, \sin t\}$.

Exercise 2.2. Find a basis for the subspace $W \subseteq P_5$ defined by

$$W = \{p \text{ in } P_5 : p(-1) = p(0) = p(1) = 0\}$$

using the following strategy: Write the general element of P_5 in the form $p(t) = a_0 + a_1 t + \cdots + a_5 t^5$, use each of the given conditions to impose a linear equation on the coefficients, and use row reduction to solve the resulting system of equations.

Show that $q(t) = t - t^3$ divides each basis element.

Exercise 2.3. Let $a < b < c$ be real numbers, and consider the quadratic *Lagrange interpolation polynomials*

$$\begin{aligned} e_1(t) &= (t - b)(t - c)/\big[(a - b)(a - c)\big], \\ e_2(t) &= (t - a)(t - c)/\big[(b - a)(b - c)\big], \\ e_3(t) &= (t - a)(t - a)/\big[(c - a)(c - b)\big]. \end{aligned}$$

(a) Evaluate each polynomial at $t = a$, b, and c.

(b) Show $(e_i)_{i=1}^3$ is a basis of P_2. Hint: Use part (a).

(c) If p is an arbitrary quadratic polynomial, find a formula (in terms of a, b, c, and p) expressing p as a linear combination from $(e_i)_{i=1}^3$, and write each of $p_0(t) = 1$, $p_1(t) = t$, and $p_2(t) = t^2$ as a linear combination from $(e_i)_{i=1}^3$. Hint: Use part (a).

Exercise 2.4. Let $V = \mathcal{F}(\mathbf{R}, \mathbf{R})$ be the vector space of real-valued functions on $\mathbf{R}$ under ordinary addition and scalar multiplication.

(a) A function ϕ is *even* if $\phi(-t) = \phi(t)$ for all t. Prove that the set $\mathcal{E}$ of even functions is a subspace of V.

(b) A function ϕ is *odd* if $\phi(-t) = -\phi(t)$ for all t. Prove that the set $\mathcal{O}$ of odd functions is a subspace of V.

(c) Show that $V = \mathcal{E} + \mathcal{O}$, and determine whether or not the sum is direct. Hint: If f is an arbitrary function, consider the functions

$$f_{\text{even}}(t) = \tfrac{1}{2}\big(f(t) + f(-t)\big), \qquad f_{\text{odd}}(t) = \tfrac{1}{2}\big(f(t) - f(-t)\big).$$

(d) Write $f(t) = e^t$ as the sum of an even and an odd function.

Exercise 2.5. Let $(V, +, \cdot)$ be a vector space, $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ elements of V.

(a) Show $\{\boldsymbol{x}, \boldsymbol{y}\}$ is linearly independent if and only if $\{\boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}-\boldsymbol{y}\}$ is linearly independent.

(b) True or false (with proof): $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ is linearly independent if and only if $\{\boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}+\boldsymbol{z}\}$ is linearly independent.

(c) True or false (with proof): $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ is linearly independent if and only if $\{\boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}-\boldsymbol{z}, \boldsymbol{y}-\boldsymbol{z}\}$ is linearly independent.

Exercise 2.6. In $\mathbf{R}^6$, give specific examples of three-dimensional subspaces W_1 and W_2 whose intersection has dimension (a) two; (b) one; (c) zero.

Exercise 2.7. If $A = [A^i_j]$ is a 2×2 matrix, the *trace* of A is defined to be $\operatorname{tr}(A) = A^1_1 + A^2_2$, the sum of the diagonal entries.

(a) Show that the set W of 2×2 matrices of trace 0 is a subspace of $\mathbf{R}^{2\times 2}$.

(b) Following Example 2.73, find bases for W, for $\operatorname{Sym}^2 \cap W$, and for $\operatorname{Skew}^2 \cap W$.

(c) Is it true that $\operatorname{Sym}^2 + W = \mathbf{R}^{2\times 2}$? Explain.

Exercise 2.8. Determine the dimensions of the spaces of upper triangular, lower triangular, and diagonal $n \times n$ matrices, and verify the Dimension Theorem for this example.

Chapter 3

Euclidean Geometry

Vectors allow us to speak of linear combinations, but not of geometric concepts such as length, angle, and volume. In order to do geometry (particularly in $\mathbf{R}^n$), we introduce additional structure.

3.1 Inner Products

Definition 3.1. Let $(V, +, \cdot)$ be a real vector space. An *inner product* on V is a function $B : V \times V \to \mathbf{R}$ satisfying the following conditions:

(i) (Bilinearity) For all $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{y}$ in V, and all real numbers c,

$$\begin{aligned} B(c\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}) &= cB(\boldsymbol{x}_1, \boldsymbol{y}) + B(\boldsymbol{x}_2, \boldsymbol{y}), \\ B(\boldsymbol{y}, c\boldsymbol{x}_1 + \boldsymbol{x}_2) &= cB(\boldsymbol{y}, \boldsymbol{x}_1) + B(\boldsymbol{y}, \boldsymbol{x}_2). \end{aligned}$$

(ii) (Symmetry) For all $\boldsymbol{x}$ and $\boldsymbol{y}$ in V, $B(\boldsymbol{y}, \boldsymbol{x}) = B(\boldsymbol{x}, \boldsymbol{y})$.

(iii) (Positive-definiteness) For all $\boldsymbol{x}$ in V, $B(\boldsymbol{x}, \boldsymbol{x}) \geq 0$, with equality if and only if $\boldsymbol{x} = \mathbf{0}^V$.

Remark 3.2. Bilinearity formalizes a two-sided "distributive law", allowing inner products of linear combinations to be expanded in terms of inner products of vectors in the combination.

To establish bilinearity for a particular function B, it suffices to verify symmetry and the first "bilinearity" condition.

Example 3.3. (The Euclidean dot product) On $(\mathbf{R}^n, +, \cdot)$, define

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{y} = x^1 y^1 + \cdots + x^n y^n = \sum_{i=1}^{n} x^i y^i.$$

AMS Open Math Notes: Works in Progress; Reference # OMN:201801.110759; Last Revised: 2018-01-20 09:27:31

The expression $\boldsymbol{x}^\mathsf{T}\boldsymbol{y}$ is the 1×1 matrix obtained by multiplying the transpose of $\boldsymbol{x}$ (a row) with $\boldsymbol{y}$ (a column), viewed as a real number.

Bilinearity amounts to the distributive law: If $\boldsymbol{x}_1 = [x_1^i]$, $\boldsymbol{x}_2 = [x_2^i]$, and $\boldsymbol{y} = [y^i]$ are arbitrary vectors and c is real, then

$$\begin{aligned}\langle c\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}\rangle &= \sum_{i=1}^n (cx_1^i + x_2^i)y^i = \sum_{i=1}^n (cx_1^i y^i + x_2^i y^i) \\ &= c\sum_{i=1}^n x_1^i y^i + \sum_{i=1}^n x_2^i y^i = c\,\langle \boldsymbol{x}_1, \boldsymbol{y}\rangle + \langle \boldsymbol{x}_2, \boldsymbol{y}\rangle\,.\end{aligned}$$

Symmetry and positive-definiteness are clear.

Example 3.4. Let $a < b$ be real. On $V = \mathcal{C}\big([a,b], \mathbf{R}\big)$, the vector space of continuous, real-valued functions on the interval $[a,b]$ under function addition and scalar multiplication, the function

$$B(f,g) = \frac{1}{b-a}\int_a^b f(t)g(t)\,dt$$

defines an inner product. Bilinearity amounts to the distributive law plus linearity of the integral; symmetry is clear; positive definiteness depends on the formal definition of continuity, and is omitted.

For concreteness, the theorems below are stated only for the dot product on $\mathbf{R}^n$. The proofs, however, use nothing more than the axioms for an inner product, and so they hold (with suitable interpretations) for an arbitrary inner product.

Magnitude

Definition 3.5. Let $\boldsymbol{x} = (x^1, \dots, x^n)$ be a vector in $\mathbf{R}^n$. The *magnitude* of $\boldsymbol{x}$ is the non-negative real number

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = \sqrt{(x^1)^2 + \cdots + (x^n)^2} = \left(\sum_{i=1}^n (x^i)^2\right)^{1/2}.$$

A vector of magnitude 1 is a *unit vector.*

Proposition 3.6. *If $\boldsymbol{x}$ is a vector in $\mathbf{R}^n$ and c is real, $\|c\boldsymbol{x}\| = |c|\,\|\boldsymbol{x}\|$. If $\boldsymbol{x} \neq \mathbf{0}^n$, then $\boldsymbol{x}$ may be written uniquely as a positive scalar multiple of a unit vector:*

$$\boldsymbol{x} = \|\boldsymbol{x}\|\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right).$$

Proof. The definition of magnitude, together with bilinearity, gives

$$\|c\boldsymbol{x}\| = \sqrt{\langle c\boldsymbol{x}, c\boldsymbol{x}\rangle} = \sqrt{c^2 \langle \boldsymbol{x}, \boldsymbol{x}\rangle} = |c|\,\|\boldsymbol{x}\|.$$

If $\boldsymbol{x} \neq \mathbf{0}^n$, a scalar multiple of $\boldsymbol{x}$ is a unit vector if and only if $\|c\boldsymbol{x}\| = 1$, if and only if $|c| = 1/\|\boldsymbol{x}\|$, i.e., $c = \pm 1/\|\boldsymbol{x}\|$. The only positive choice is $c = 1/\|\boldsymbol{x}\|$. □

Definition 3.7. If $\boldsymbol{x} \neq \mathbf{0}$, the unit vector $\boldsymbol{x}/\|\boldsymbol{x}\|$ is obtained by *normalizing* $\boldsymbol{x}$.

Remark 3.8. A unit vector may be viewed as a "pure direction". The proposition gives precise meaning to the "principle" that a vector is a quantity having both magnitude and direction.

Theorem 3.9 (Cauchy-Schwarz Inequality)**.** *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors in $\mathbf{R}^n$, then*

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|,$$

with equality if and only if $\boldsymbol{x}$ and $\boldsymbol{y}$ are proportional.

Proof. If $\boldsymbol{x} = \mathbf{0}$, then $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$ and the conclusion of the theorem is immediate. Otherwise, for each real number t, consider the vector $t\boldsymbol{x} + \boldsymbol{y}$. Expanding $\|t\boldsymbol{x} + \boldsymbol{y}\|^2 = \langle t\boldsymbol{x} + \boldsymbol{y}, t\boldsymbol{x} + \boldsymbol{y}\rangle$ as a function of t,

$$\begin{aligned} 0 \leq \langle t\boldsymbol{x} + \boldsymbol{y}, t\boldsymbol{x} + \boldsymbol{y}\rangle &= t^2 \langle \boldsymbol{x}, \boldsymbol{x}\rangle + 2t\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle \\ &= t^2\|\boldsymbol{x}\|^2 + 2t\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2. \end{aligned}$$

In other words, $0 \leq At^2 + Bt + C$, with $A = \|\boldsymbol{x}\|^2$, $B = 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle$, and $C = \|\boldsymbol{y}\|^2$.

A non-negative quadratic function satisfies $B^2 - 4AC \leq 0$, or $B^2 \leq 4AC$; otherwise the quadratic formula would give two distinct real roots, and the quadratic would be negative in between. Plugging in the definitions of A, B, and C gives

$$(\langle \boldsymbol{x}, \boldsymbol{y}\rangle)^2 \leq \|\boldsymbol{x}\|^2\,\|\boldsymbol{y}\|^2, \qquad \text{or} \quad |\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|.$$

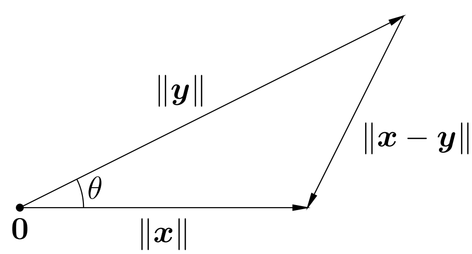
Equality holds if and only if the quadratic has a double root, if and only if there exists a real number t such that $t\boldsymbol{x} + \boldsymbol{y} = \mathbf{0}$, if and only if $\boldsymbol{x}$ and $\boldsymbol{y}$ are proportional. □

Angle

Theorem 3.10. *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors in $\mathbf{R}^n$, then the angle θ between $\boldsymbol{x}$ and $\boldsymbol{y}$ satisfies*

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \cos\theta.$$

Proof. If $\boldsymbol{x} = \mathbf{0}$ or $\boldsymbol{y} = \mathbf{0}$, the equality in the theorem is immediate. It remains to handle the case with $\boldsymbol{x}$ and $\boldsymbol{y}$ both non-zero. Let θ be the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ in the plane containing these two vectors. The



triangle with vertices $\mathbf{0}$, $\boldsymbol{x}$, and $\boldsymbol{y}$ has sides of magnitude $\|\boldsymbol{x}\|$, $\|\boldsymbol{y}\|$, and $\|\boldsymbol{x}-\boldsymbol{y}\|$. By the Law of Cosines,

$$\|\boldsymbol{x}-\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \cos\theta.$$

But the left-hand side can be expanded as

$$\langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}-\boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle .$$

Equating and canceling gives the theorem. ☐

Corollary 3.11. *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are* non-zero *vectors, there is a unique real number θ in the interval $[0,\pi]$ such that*

$$\cos\theta = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \, \|\boldsymbol{y}\|}.$$

Remark 3.12. The expression on the right lies in the interval $[-1,1]$ by the Cauchy-Schwarz inequality.

Definition 3.13. Two vectors $\boldsymbol{x}$, $\boldsymbol{y}$ in $\mathbf{R}^n$ are *orthogonal* if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

Example 3.14. The vectors $\boldsymbol{x} = (1,1,1)$ and $\boldsymbol{y} = (1,-2,1)$ in $\mathbf{R}^3$ are orthogonal.

Example 3.15. If $\boldsymbol{x} = (1, 0)$ and $\boldsymbol{y} = (1, 1)$, then

$$\|\boldsymbol{x}\|^2 = \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 1, \qquad \|\boldsymbol{y}\|^2 = \langle \boldsymbol{y}, \boldsymbol{y} \rangle = 2, \qquad \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 1.$$

Consequently, $\cos\theta = 1/\sqrt{2}$, so $\theta = \pi/4$. This agrees with geometric expectation, since $\boldsymbol{x}$ is a side and $\boldsymbol{y}$ a diagonal of a square.

We can perform angle computations for vectors we cannot easily visualize. If $\boldsymbol{x} = (1, 0, 0, 0)$ and $\boldsymbol{y} = (1, 1, 1, 1)$, then

$$\|\boldsymbol{x}\|^2 = 1, \qquad \|\boldsymbol{y}\|^2 = 4, \qquad \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 1.$$

Consequently, $\cos\theta = 1/2$, so $\theta = \pi/3$. (Generally, the angle between a standard basis vector and the "diagonal of an n-cube" is not a "nice" multiple of π.)

Theorem 3.16 (The Pythagorean Theorem in $\mathbf{R}^n$)**.** *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors in $\mathbf{R}^n$, then*

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 \quad \text{if and only if } \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are orthogonal.}$$

Proof. Expanding the left-hand side as a dot product,

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle .$$

The theorem follows immediately. □

3.2 Orthogonal Projection

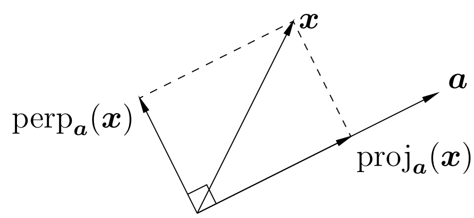
Definition 3.17. Let $\boldsymbol{a}$ be a *non-zero* vector in $\mathbf{R}^n$. If $\boldsymbol{x} \in \mathbf{R}^n$, a pair of vectors $\boldsymbol{x}_{\boldsymbol{a}}$, $\boldsymbol{x}_{\boldsymbol{a}^\perp}$ is called a *decomposition* of $\boldsymbol{x}$ into *parallel* and *orthogonal* components with respect to $\boldsymbol{a}$ if (i) $\boldsymbol{x} = \boldsymbol{x}_{\boldsymbol{a}} + \boldsymbol{x}_{\boldsymbol{a}^\perp}$; (ii) $\boldsymbol{x}_{\boldsymbol{a}} = c\boldsymbol{a}$ for some real number c; (iii) $\langle \boldsymbol{x}_{\boldsymbol{a}^\perp}, \boldsymbol{a} \rangle = 0$.

Theorem 3.18. *Let $\boldsymbol{a}$ be a non-zero vector in $\mathbf{R}^n$. For each $\boldsymbol{x}$ in $\mathbf{R}^n$, there is a unique decomposition of $\boldsymbol{x}$ into parallel and orthogonal components with respect to $\boldsymbol{a}$, given by the formulas*

$$\mathrm{proj}_{\boldsymbol{a}}(\boldsymbol{x}) = \frac{\langle \boldsymbol{x}, \boldsymbol{a} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle} \cdot \boldsymbol{a}, \qquad \mathrm{perp}_{\boldsymbol{a}}(\boldsymbol{x}) = \boldsymbol{x} - \mathrm{proj}_{\boldsymbol{a}}(\boldsymbol{x}).$$

Proof. Let $\boldsymbol{x}$ be an arbitrary vector in $\mathbf{R}^n$. If c is a real number, the difference $\boldsymbol{x} - c\boldsymbol{a}$ is orthogonal to $\boldsymbol{a}$ if and only if c satisfies the equation $0 = \langle \boldsymbol{x} - c\boldsymbol{a}, \boldsymbol{a} \rangle = \langle \boldsymbol{x}, \boldsymbol{a} \rangle - c\langle \boldsymbol{a}, \boldsymbol{a} \rangle$. By elementary algebra, $c = \langle \boldsymbol{x}, \boldsymbol{a} \rangle / \langle \boldsymbol{a}, \boldsymbol{a} \rangle$. □

Definition 3.19. If $\boldsymbol{a}$ is a non-zero vector in $\mathbf{R}^n$, then $\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x})$ is called the (orthogonal) *projection* of $\boldsymbol{x}$ on $\boldsymbol{a}$.



Remark 3.20. If $\|\boldsymbol{u}\| = 1$, then $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{u} \rangle \cdot \boldsymbol{u}$.

Example 3.21. Let $\boldsymbol{a} = (2, 1)$, and $\boldsymbol{x} = (x^1, x^2)$ an arbitrary vector in $\mathbf{R}^2$. We have $\langle \boldsymbol{a}, \boldsymbol{a} \rangle = 5$, $\langle \boldsymbol{x}, \boldsymbol{a} \rangle = 2x^1 + x^2$, and

$$\begin{aligned}
\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x}) &= \frac{2x^1 + x^2}{5}(2, 1) = \tfrac{1}{5}(4x^1 + 2x^2, 2x^1 + x^2),\\
\operatorname{perp}_{\boldsymbol{a}}(\boldsymbol{x}) &= \boldsymbol{x} - \operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x}) = \tfrac{1}{5}(x^1 - 2x^2, -2x^1 + 4x^2).
\end{aligned}$$

As expected, $\langle \operatorname{perp}_{\boldsymbol{a}}(\boldsymbol{x}), \boldsymbol{a} \rangle = \frac{1}{5}\big[2(x^1 - 2x^2) + (-2x^1 + 4x^2)\big] = 0$.

Example 3.22. If θ is real, $\boldsymbol{a} = (\cos\theta, \sin\theta)$, and $\boldsymbol{a}^\perp = (-\sin\theta, \cos\theta)$, then

$$\begin{aligned}
\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x}) &= (x^1\cos\theta + x^2\sin\theta)(\cos\theta, \sin\theta),\\
\operatorname{proj}_{\boldsymbol{a}^\perp}(\boldsymbol{x}) &= (x^2\cos\theta - x^1\sin\theta)(-\sin\theta, \cos\theta) = \boldsymbol{a} - \operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x}).
\end{aligned}$$

Example 3.23. If $\boldsymbol{a} = (1, 1, 1, 1)$, then $\langle \boldsymbol{a}, \boldsymbol{a} \rangle = 4$. For an arbitrary vector $\boldsymbol{x} = (x^1, x^2, x^3, x^4)$ in $\mathbf{R}^4$ we have $\langle \boldsymbol{x}, \boldsymbol{a} \rangle = x^1 + x^2 + x^3 + x^4$,

$$\begin{aligned}
\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x}) &= \frac{x^1 + x^2 + x^3 + x^4}{4}(1, 1, 1, 1),\\
\operatorname{perp}_{\boldsymbol{a}}(\boldsymbol{x}) &= \boldsymbol{x} - \operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x}).
\end{aligned}$$

Orthonormal Bases and Orthogonal Matrices

Definition 3.24. A set $(\boldsymbol{u}_j)_{j=1}^k$ in $\mathbf{R}^n$ is *orthonormal* if each element is a unit vector and any two elements are orthogonal, i.e., $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \delta_i^j$.

Theorem 3.25. *If $S = (\boldsymbol{u}_j)_{j=1}^k$ is an orthonormal set of vectors in $\mathbf{R}^n$, and if $\boldsymbol{x} = \sum_j x^j \boldsymbol{u}_j$ for some scalars $(x^j)_{j=1}^k$, then $x^j = \langle \boldsymbol{x}, \boldsymbol{u}_j \rangle$. In particular, S is linearly independent.*

Proof. If $\boldsymbol{x} = \sum_j x^j \boldsymbol{u}_j$, then for each i,

$$\langle \boldsymbol{x}, \boldsymbol{u}_i \rangle = \Big\langle \sum\nolimits_j x^j \boldsymbol{u}_j, \boldsymbol{u}_i \Big\rangle = \sum_{j=1}^k x^j \langle \boldsymbol{u}_j, \boldsymbol{u}_i \rangle = \sum_{j=1}^k x^j \delta_j^i = x^i.$$

If $\boldsymbol{x} = \mathbf{0}^n$ is a linear combination from S, then $x^j = \langle \boldsymbol{x}, \boldsymbol{u}_j \rangle = 0$ for all j, so S is linearly independent. □

Corollary 3.26. *If $S = (\boldsymbol{u}_j)_{j=1}^n$ is an orthonormal set in $\mathbf{R}^n$, then S is a basis, and for every $\boldsymbol{x}$ in $\mathbf{R}^n$,*

$$\boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{u}_1 \rangle \cdot \boldsymbol{u}_1 + \cdots + \langle \boldsymbol{x}, \boldsymbol{u}_n \rangle \cdot \boldsymbol{u}_n = \sum_{j=1}^n \langle \boldsymbol{x}, \boldsymbol{u}_j \rangle \cdot \boldsymbol{u}_j.$$

Proof. Since S is a linearly independent set of n elements in $\mathbf{R}^n$ by Theorem 3.25, S is a basis by Corollary 2.68. It follows that an arbitrary vector $\boldsymbol{x}$ may be written as a linear combination from S. Theorem 3.25 gives the form of the coefficients. □

Corollary 3.27. *If $(\boldsymbol{u}_j)_{j=1}^n$ is orthonormal in $\mathbf{R}^n$, then $\sum_j \boldsymbol{u}_j \boldsymbol{u}_j^\mathsf{T} = I_n$.*

Proof. For every $\boldsymbol{x}$ in $\mathbf{R}^n$, $\boldsymbol{u}_j^\mathsf{T} \boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{u}_j \rangle$. By Corollary 3.26,

$$\left(\sum_{j=1}^n \boldsymbol{u}_j \boldsymbol{u}_j^\mathsf{T} \right) \boldsymbol{x} = \sum_{j=1}^n \boldsymbol{u}_j (\boldsymbol{u}_j^\mathsf{T} \boldsymbol{x}) = \sum_{j=1}^n \langle \boldsymbol{x}, \boldsymbol{u}_j \rangle \cdot \boldsymbol{u}_j = \boldsymbol{x} = I_n \boldsymbol{x}$$

for all $\boldsymbol{x}$, so $\sum_j \boldsymbol{u}_j \boldsymbol{u}_j^\mathsf{T} = I_n$ by Corollary 1.23. □

Definition 3.28. An $n \times n$ real matrix P is *orthogonal* if $P^{-1} = P^\mathsf{T}$, i.e., if $P^\mathsf{T} P = I_n$ and $P P^\mathsf{T} = I_n$.

Example 3.29. Let θ be real. The following matrices are orthogonal:

$$\mathrm{Rot}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \qquad \mathrm{Ref}_{\theta/2} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

(The notation is explained in Examples 4.17 and 4.19.)

Example 3.30. A permutation matrix (Exercise 1.13) is orthogonal.

Theorem 3.31. *The following are equivalent for an $n \times n$ real matrix P:*

(i) *The columns of P form an orthonormal set in $\mathbf{R}^n$.*

(ii) *The columns of P^T form an orthonormal set in $\mathbf{R}^n$.*

(iii) *P is orthogonal.*

Proof. In general, if A and B are $n\times n$ matrices, the (i,j) entry of $A^{\mathsf{T}}B$ is the dot product of the ith column of A and the jth column of B. Similarly, the (i,j) entry of AB^{T} is the dot product of the ith column of A^{T} and the jth column of B^{T}.

Since $I_n = [\delta^i_j]$, the equation $P^{\mathsf{T}}P = I_n$ holds if and only if the columns of P are orthonormal, and $PP^{\mathsf{T}} = I_n$ holds if and only if the columns of P^{T} are orthonormal. The content of the proof is to show each of these equations implies the other.

((i) if and only if (ii)). If the columns of P are orthonormal, then $P^{\mathsf{T}}P = I_n$. Further, the columns of P are a basis of $\mathbf{R}^n$ by Theorem 3.25. Consequently, the reduced row-echelon form of P^{T} is I_n, so P^{T} is invertible by Theorem 1.49. By Theorem 1.30 (ii), $PP^{\mathsf{T}} = I_n$ as well, so the columns of P^{T} are orthonormal.

The converse implication follows *mutatis mutandis* by exchanging the roles of P and P^{T}.

Since (iii) is equivalent to "$P^{\mathsf{T}}P = I_n$ and $PP^{\mathsf{T}} = I_n$", the proof is complete. □

Theorem 3.32. *The set $O(n)$ of $n\times n$ real orthogonal matrices is a group under matrix multiplication.*

Proof. The identity matrix I_n is orthogonal, and acts as the multiplicative identity element in $O(n)$.

If P and Q are orthogonal, then

$$(PQ)^{\mathsf{T}} = Q^{\mathsf{T}}P^{\mathsf{T}} = Q^{-1}P^{-1} = (PQ)^{-1},$$

so PQ is orthogonal. That is, $O(n)$ is closed under multiplication.

By Theorem 3.31, P is orthogonal if and only if $P^{\mathsf{T}} = P^{-1}$ is orthogonal, so $O(n)$ is closed under inversion. □

Theorem 3.33. *The following are equivalent for an $n\times n$ real matrix P:*

(i) $\langle P\boldsymbol{x}, P\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ *for all* $\boldsymbol{x}$, $\boldsymbol{y}$ *in* $\mathbf{R}^n$.

(ii) P *is orthogonal.*

Proof. Since $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}$, we have

$$\langle P\boldsymbol{x}, P\boldsymbol{y}\rangle = (P\boldsymbol{x})^{\mathsf{T}}(P\boldsymbol{y}) = (\boldsymbol{x}^{\mathsf{T}}P^{\mathsf{T}})(P\boldsymbol{y}) = \boldsymbol{x}^{\mathsf{T}}(P^{\mathsf{T}}P)\boldsymbol{y},$$

and therefore

$$\langle P\boldsymbol{x}, P\boldsymbol{y}\rangle - \langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^{\mathsf{T}}(P^{\mathsf{T}}P)\boldsymbol{y} - \boldsymbol{x}^{\mathsf{T}}(I_n)\boldsymbol{y} = \boldsymbol{x}^{\mathsf{T}}(P^{\mathsf{T}}P - I_n)\boldsymbol{y}$$

for all $\boldsymbol{x}$, $\boldsymbol{y}$. It follows that $\langle P\boldsymbol{x}, P\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}$, $\boldsymbol{y}$ in $\mathbf{R}^n$ if and only if $P^{\mathsf{T}}P - I_n = \mathbf{0}^{n\times n}$, if and only if P is orthogonal. □

The Gram-Schmidt Algorithm

Theorem 3.34. *If W is a subspace of $(\mathbf{R}^n, +, \cdot)$, then W has an orthonormal basis.*

Proof. The idea is to start from an arbitrary basis $(\boldsymbol{v}_j)_{j=1}^m$ of W and "orthonormalize" the vectors inductively, constructing, for $k = 1, \dots, m$, an orthonormal basis $(\boldsymbol{u}_j)_{j=1}^k$ of the space $W_k = \mathrm{Span}(\boldsymbol{v}_j)_{j=1}^k$, see Figure 3.1. The process is called the *Gram-Schmidt* algorithm. When $k = m$, we obtain an orthonormal basis of W.

Since $\boldsymbol{v}_1 \neq \mathbf{0}^n$, we may set $\boldsymbol{u}_1 = \boldsymbol{v}_1/\|\boldsymbol{v}_1\|$. Assume inductively that, for some $k \geq 1$, we have constructed an orthonormal basis $(\boldsymbol{u}_j)_{j=1}^k$ of the space $W_k = \mathrm{Span}(\boldsymbol{v}_j)_{j=1}^k$. Define

$$\boldsymbol{u}'_{k+1} = \boldsymbol{v}_{k+1} - \sum_{j=1}^{k} \langle \boldsymbol{v}_{k+1}, \boldsymbol{u}_j \rangle \cdot \boldsymbol{u}_j.$$

Note that $\boldsymbol{u}'_{k+1} \neq \mathbf{0}^n$ since $\boldsymbol{v}_{k+1}$ does not lie in $W_k = \mathrm{Span}(\boldsymbol{u}_j)_{j=1}^k$, and that $\left\langle \boldsymbol{u}'_{k+1}, \boldsymbol{u}_j \right\rangle = 0$ for $j = 1, \dots, k$ by direct computation. We may therefore put $\boldsymbol{u}_{k+1} = \boldsymbol{u}'_{k+1}/\|\boldsymbol{u}'_{k+1}\|$.

Since each element of the linearly independent set $(\boldsymbol{u}_j)_{j=1}^{k+1}$ is a linear combination from $(\boldsymbol{v}_j)_{j=1}^{k+1}$, each set is a basis of W_{k+1}. □

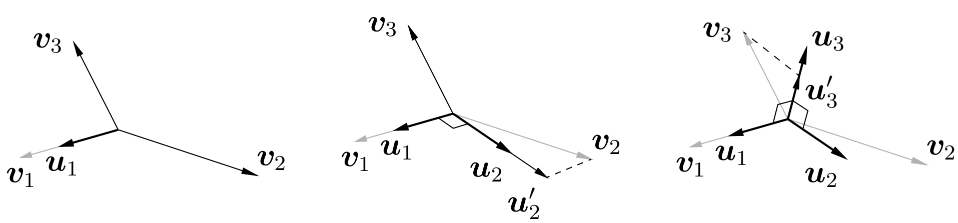


Figure 3.1: The Gram-Schmidt algorithm.

The Orthogonal Complement of a Subspace

Definition 3.35. If W is a subspace of $(\mathbf{R}^n, +, \cdot)$ equipped with the Euclidean inner product, the *orthogonal complement* $W^\perp$ is

$$W^\perp = \{\boldsymbol{x} \text{ in } \mathbf{R}^n : \langle \boldsymbol{x}, \boldsymbol{w} \rangle = 0 \text{ for all } \boldsymbol{w} \text{ in } W\}.$$

Remark 3.36. The fact that $W^\perp$ *is* a subspace comes from bilinearity of the dot product: If $\boldsymbol{x}$, $\boldsymbol{y}$ are in $W^\perp$ and c is real, then

$$\langle c\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{w}\rangle = c\,\langle \boldsymbol{x}, \boldsymbol{w}\rangle + \langle \boldsymbol{y}, \boldsymbol{w}\rangle = 0 \text{ for all } \boldsymbol{w} \text{ in } W.$$

Theorem 3.37. *Let W be a subspace of $(\mathbf{R}^n, +, \cdot)$ and $(\boldsymbol{u}_j)_{j=1}^m$ an orthonormal basis of W. If $\boldsymbol{x}$ is an arbitrary vector in $\mathbf{R}^n$, define*

$$\operatorname{proj}_W(\boldsymbol{x}) = \sum_{j=1}^{m} \langle \boldsymbol{x}, \boldsymbol{u}_j\rangle \cdot \boldsymbol{u}_j.$$

(i) *We have* $\operatorname{proj}_W(\boldsymbol{x}) \in W$ *and* $\boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x}) \in W^\perp$.

(ii) *For every $\boldsymbol{w}$ in W, we have $\|\boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x})\| \leq \|\boldsymbol{x} - \boldsymbol{w}\|$, with equality if and only if $\boldsymbol{w} = \operatorname{proj}_W(\boldsymbol{x})$.*

Proof. (i). Since $S = (\boldsymbol{u}_j)_{j=1}^m \subseteq W$ and $\operatorname{proj}_W(\boldsymbol{x})$ is a linear combination from S, $\operatorname{proj}_W(\boldsymbol{x}) \in W$.

The proof of Theorem 3.25 shows that $\langle \operatorname{proj}_W(\boldsymbol{x}), \boldsymbol{u}_j\rangle = \langle \boldsymbol{x}, \boldsymbol{u}_j\rangle$ for $j = 1, \dots, m$. It follows that $\langle \boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x}), \boldsymbol{u}_j\rangle = 0$ for $j = 1, \dots, m$. Since each element of W is a linear combination from $(\boldsymbol{u}_j)_{j=1}^m$, $\langle \boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x}), \boldsymbol{w}\rangle = 0$ for all $\boldsymbol{w}$ in W, i.e., $\boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x}) \in W^\perp$.

(ii). Let $\boldsymbol{w}$ be an arbitrary vector in W. Since $\operatorname{proj}_W(\boldsymbol{x}) - \boldsymbol{w}$ is in W (as a difference of vectors in W) and therefore orthogonal to $\boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x})$ by part (i), the Pythagorean Theorem gives

$$\begin{aligned}
\|\boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x})\|^2 &\leq \|\boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x})\|^2 + \|\operatorname{proj}_W(\boldsymbol{x}) - \boldsymbol{w}\|^2 \\
&= \|(\boldsymbol{x} - \operatorname{proj}_W(\boldsymbol{x})) + (\operatorname{proj}_W(\boldsymbol{x}) - \boldsymbol{w})\|^2 \\
&= \|\boldsymbol{x} - \boldsymbol{w}\|^2,
\end{aligned}$$

with equality if and only if $\boldsymbol{w} = \operatorname{proj}_W(\boldsymbol{x})$. □

Remark 3.38. Though the *definition* of $\operatorname{proj}_W(\boldsymbol{x})$ depended on a choice of orthonormal basis of W, the *value* does not: Property (ii) shows there is at most one element of W that is closer to $\boldsymbol{x}$ than all other elements of W, and $\operatorname{proj}_W(\boldsymbol{x})$ satisfies this “best approximation property”.

Corollary 3.39. *If $W \subseteq \mathbf{R}^n$ is a subspace, then $\mathbf{R}^n = W \oplus W^\perp$.*

Proof. ($W \cap W^\perp = \{\mathbf{0}^n\}$). If $\boldsymbol{x} \in W \cap W^\perp$, then $\boldsymbol{x}$ is orthogonal to itself. That is, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$. By positive-definiteness, $\boldsymbol{x} = \mathbf{0}^n$.

($\mathbf{R}^n = W + W^\perp$). By the Gram-Schmidt algorithm, W has an orthonormal basis $(\boldsymbol{u}_j)_{j=1}^m$. If $\boldsymbol{x} \in \mathbf{R}^n$, then by Theorem 3.37,

$$\boldsymbol{x} = \mathrm{proj}_W(\boldsymbol{x}) + \big(\boldsymbol{x} - \mathrm{proj}_W(\boldsymbol{x})\big) \in W + W^\perp. \qquad \square$$

Corollary 3.40. *If W is a k-dimensional subspace of $(\mathbf{R}^n, +, \cdot)$, there exists a homogeneous linear system $A\boldsymbol{x} = \mathbf{0}^{n-k}$ of $(n-k)$ equations in n variables whose solution set is precisely W.*

Proof. A linear equation $a_1 x^1 + \cdots + a_n x^n = 0$ may be interpreted as asserting that the product of the row matrix $\boldsymbol{a}^\mathsf{T} = [a_j]$ with the column $\boldsymbol{x} = [x^j]$ is 0, i.e., that the vectors $\boldsymbol{a} = [a^j]$ and $\boldsymbol{x}$ are orthogonal.

Let $W \subseteq \mathbf{R}^n$ be a subspace. By Corollary 3.39, the orthogonal complement $W^\perp$ has dimension $(n-k)$. Pick a basis $(\boldsymbol{a}_i)_{i=1}^{n-k}$, with $\boldsymbol{a}_i = [a_i^j]$, and form the matrix $A = [a_j^i]$ whose ith row is $\boldsymbol{a}_i^\mathsf{T}$.

If $\boldsymbol{x} \in \mathbf{R}^n$, then $A\boldsymbol{x} = \mathbf{0}^{n-k}$ if and only if $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle = 0$ for each $i = 1, \dots, n-k$, if and only if $\boldsymbol{x} \in (W^\perp)^\perp$, if and only if $\boldsymbol{x} \in W$ by Corollary 3.39. $\square$

Example 3.41. Let $\boldsymbol{v}_0 = (1, 1, 1, 1)$, and $W = \mathrm{Span}(\boldsymbol{v}_0)^\perp \subseteq \mathbf{R}^4$. Find an orthonormal basis of W.

A vector $\boldsymbol{x} = (x^1, x^2, x^3, x^4)$ is in W if and only if

$$0 = \langle \boldsymbol{x}, \boldsymbol{v}_0 \rangle = x^1 + x^2 + x^3 + x^4.$$

The coefficient matrix $\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$ is already in reduced row-echelon form. The variables x^2, x^3, and x^4 are free, and the general solution is

$$\begin{bmatrix} -(x^2 + x^3 + x^4) \\ x^2 \\ x^3 \\ x^4 \end{bmatrix} = x^2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x^3 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x^4 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Applying Gram-Schmidt to the columns $(\boldsymbol{v}_j)_{j=1}^3$ on the right-hand side,

$$\boldsymbol{u}_1 = \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Next, since $\langle \boldsymbol{v}_2, \boldsymbol{u}_1 \rangle = 1/\sqrt{2}$,

$$\boldsymbol{v}_2' = \boldsymbol{v}_2 - \langle \boldsymbol{v}_2, \boldsymbol{u}_1 \rangle \cdot \boldsymbol{u}_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \tfrac{1}{2} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix} = \tfrac{1}{2} \begin{bmatrix} -1 \\ -1 \\ 2 \\ 0 \end{bmatrix};$$

normalizing gives $\boldsymbol{u}_2 = \dfrac{\boldsymbol{v}_2'}{\|\boldsymbol{v}_2'\|} = \dfrac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ -1 \\ 2 \\ 0 \end{bmatrix}$.

Finally,

$$\begin{aligned} \boldsymbol{v}_3' &= \boldsymbol{v}_3 - \langle \boldsymbol{v}_3, \boldsymbol{u}_1 \rangle \cdot \boldsymbol{u}_1 - \langle \boldsymbol{v}_3, \boldsymbol{u}_2 \rangle \cdot \boldsymbol{u}_2 \\ &= \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \tfrac{1}{2} \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \end{bmatrix} - \tfrac{1}{6} \begin{bmatrix} -1 \\ -1 \\ 2 \\ 0 \end{bmatrix} = \tfrac{1}{3} \begin{bmatrix} -1 \\ -1 \\ -1 \\ 3 \end{bmatrix}, \end{aligned}$$

so $\boldsymbol{u}_3 = \dfrac{\boldsymbol{v}_3'}{\|\boldsymbol{v}_3'\|} = \dfrac{1}{2\sqrt{3}} \begin{bmatrix} -1 \\ -1 \\ -1 \\ 3 \end{bmatrix}$.

Remark 3.42. Normalizing a vector is equivalent to normalizing a non-zero scalar multiple; in the preceding example, we drop the fraction multipliers before computing magnitudes.

3.3 The Determinant

Definition 3.43. Let $(\boldsymbol{v}_j)_{j=1}^n$ be a set of vectors in $\mathbf{R}^n$. The *box* (or *parallelipiped*) they span is the set of linear combinations

$$\sum_{j=1}^{n} x^j \boldsymbol{v}_j, \quad 0 \leq x^j \leq 1 \text{ for } j = 1, \ldots, n.$$

The vector $\boldsymbol{v}_j$ is the jth *edge* of the box.

The *unit cube* is the box spanned by the standard basis $(\boldsymbol{e}_j)_{j=1}^n$.

We wish to define a notion of "oriented (n-dimensional) volume" for a box together with an ordering of its edges, generalizing oriented

area of a plane parallelogram, Figure 3.2. Our approach is to give *axioms* for a real-valued function on the set $(\mathbf{R}^n)^n$ of ordered n-tuples of vectors in $\mathbf{R}^n$ in such a way that oriented volume can be computed by assembling $(\boldsymbol{v}_j)$ into an $n \times n$ matrix and using row reduction.

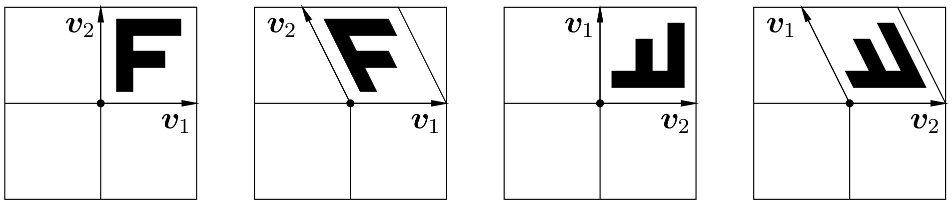


Figure 3.2: Positive and negative oriented area.

Along the way, we find a *polynomial formula* for oriented volume in terms of the components of the $(\boldsymbol{v}_j)$. The catch is, the formula contains $n!$ summands. For matrices of size 2×2 (two terms) and 3×3 (six terms) the formula is suitable for practical use. Even for 10×10 matrices ($10! = 3,628,800$ terms), however, row reduction is the only serious approach for computing oriented volume.

Theorem 3.44. *There exists a unique function* $\mathrm{vol} : (\mathbf{R}^n)^n \to \mathbf{R}$ *satisfying the following conditions for all* $(\boldsymbol{v}_j)_{j=1}^n$ *in* $\mathbf{R}^n$*:*

(i) *For all* $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ *in* $\mathbf{R}^n$ *and all real* c*,*

$$\mathrm{vol}(c\boldsymbol{x}_1+\boldsymbol{x}_2, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n) = c\,\mathrm{vol}(\boldsymbol{x}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)+\mathrm{vol}(\boldsymbol{x}_2, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n).$$

(ii) *Swapping two arguments changes the sign:*

$$\mathrm{vol}(\dots, \boldsymbol{v}_j, \dots, \boldsymbol{v}_i, \dots) = -\,\mathrm{vol}(\dots, \boldsymbol{v}_i, \dots, \boldsymbol{v}_j, \dots).$$

In particular, $\mathrm{vol}(\dots, \boldsymbol{v}_i, \dots, \boldsymbol{v}_i, \dots) = 0$.

(iii) $\mathrm{vol}(\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n) = 1$.

Remark 3.45. The "distributive law" in (i) extends to arbitrary linear combinations in the first argument by induction on the number of summands.

In conjunction with (ii), the property of *skew-symmetry*, there is a similar distributive law in the jth argument: Swap the first and jth arguments using (ii), apply (i) to break up the linear combination, then re-swap the first and jth arguments in each summand; the overall effect is to break up a linear combination in the jth argument.

Before giving a proof, we look explicitly at the cases $n = 2$ and 3.

Example 3.46. If $\boldsymbol{v}_1 = a^1\boldsymbol{e}_1 + a^2\boldsymbol{e}_2$ and $\boldsymbol{v}_2 = b^1\boldsymbol{e}_1 + b^2\boldsymbol{e}_2$, then $\operatorname{vol}(\boldsymbol{v}_1, \boldsymbol{v}_2) = a^1b^2 - a^2b^1$. Indeed, the "distributive law" gives

$$\begin{aligned}
\operatorname{vol}(\boldsymbol{v}_1, \boldsymbol{v}_2) &= \operatorname{vol}(a^1\boldsymbol{e}_1 + a^2\boldsymbol{e}_2, b^1\boldsymbol{e}_1 + b^2\boldsymbol{e}_2) \\
&= a^1b^1 \operatorname{vol}(\boldsymbol{e}_1, \boldsymbol{e}_1) + a^1b^2 \operatorname{vol}(\boldsymbol{e}_1, \boldsymbol{e}_2) \\
&\quad + a^2b^1 \operatorname{vol}(\boldsymbol{e}_2, \boldsymbol{e}_1) + a^2b^2 \operatorname{vol}(\boldsymbol{e}_2, \boldsymbol{e}_2).
\end{aligned}$$

Since $\operatorname{vol}(\boldsymbol{e}_1, \boldsymbol{e}_1) = \operatorname{vol}(\boldsymbol{e}_2, \boldsymbol{e}_2) = 0$, and $\operatorname{vol}(\boldsymbol{e}_2, \boldsymbol{e}_1) = -\operatorname{vol}(\boldsymbol{e}_1, \boldsymbol{e}_2)$ by (ii),

$$\begin{aligned}
\operatorname{vol}(\boldsymbol{v}_1, \boldsymbol{v}_2) &= a^1b^2 \operatorname{vol}(\boldsymbol{e}_1, \boldsymbol{e}_2) + a^2b^1 \operatorname{vol}(\boldsymbol{e}_2, \boldsymbol{e}_1) \\
&= (a^1b^2 - a^2b^1) \operatorname{vol}(\boldsymbol{e}_1, \boldsymbol{e}_2) \\
&= a^1b^2 - a^2b^1.
\end{aligned}$$

Conceptually, the "distributive law" allows $\operatorname{vol}(\boldsymbol{v}_1, \boldsymbol{v}_2)$ to be expressed in terms of $\operatorname{vol}(\boldsymbol{e}_i, \boldsymbol{e}_j)$ for all $2^2 = 4$ ordered pairs of indices i and j. Terms with repeated index are 0, so only $2! = 2$ terms are non-zero, one with positive sign, one with negative sign.

Example 3.47. If $\boldsymbol{v}_1 = (a^1, a^2, a^3)$, $\boldsymbol{v}_2 = (b^1, b^2, b^3)$, $\boldsymbol{v}_3 = (c^1, c^2, c^3)$, then

$$\operatorname{vol}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) = a^1(b^2c^3 - b^3c^2) + a^2(b^3c^1 - b^1c^3) + a^3(b^1c^2 - b^2c^1).$$

This can be handled by brute force, but now instead of $2^2 = 4$ terms (only two of which are non-zero), there are $3^3 = 27$ terms, of which only $3! = 6$ are non-zero. Examining the structure of the computation saves enough work to be worthwhile.

When we use $\boldsymbol{v}_1 = a^1\boldsymbol{e}_1 + a^2\boldsymbol{e}_2 + a^3\boldsymbol{e}_3$ (etc.) to expand $\operatorname{vol}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$, we get summands

$$a^ib^jc^k \operatorname{vol}(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{e}_k),$$

with each index i, j, k taking on each value 1, 2, 3. Because of skew-symmetry, any term with a repeated index is zero. Thus, there is one non-zero term for each permutation of the indices 1, 2, 3, and

$$\operatorname{vol}(\boldsymbol{e}_i, \boldsymbol{e}_j, \boldsymbol{e}_k) = \pm \operatorname{vol}(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3) = \pm 1,$$

the sign being determined by the number of "swaps" needed to put i, j, k in increasing order. The end result is

$$\begin{aligned}
\operatorname{vol}(\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3) &= a^1b^2c^3 - a^1b^3c^2 + a^2b^3c^1 - a^2b^1c^3 + a^3b^1c^2 - a^3b^2c^1 \\
&= a^1(b^2c^3 - b^3c^2) + a^2(b^3c^1 - b^1c^3) + a^3(b^1c^2 - b^2c^1).
\end{aligned}$$

Proof. (Uniqueness). Because an arbitrary vector $\boldsymbol{v}_j$ can be expressed (uniquely) as a linear combination of standard basis vectors, the "distributive law" shows vol is completely determined by its values on arbitrary n-tuples of standard basis vectors.

By (ii), if any vector is repeated as an argument, the corresponding summand is 0. That is, vol is determined by its values on n-tuples of *distinct* standard basis vectors, i.e., on permutations of the standard basis. But every ordering of the standard basis vectors can be put into "standard" order by swapping arguments. It follows that vol is completely determined by $\mathrm{vol}(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$, which is equal to 1 by (iii).

To implement the preceding conceptual calculation as a formula, write, for each $j = 1, \dots, n$,

$$\boldsymbol{v}_j = A_j^1 \boldsymbol{e}_1 + \cdots + A_j^n \boldsymbol{e}_n = \sum_{i=1}^n A_j^i \boldsymbol{e}_i.$$

Each non-zero term in $\mathrm{vol}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ comes from a permutation σ. If $(-1)^\sigma$ denotes the sign of σ, i.e., $+1$ if σ is an even permutation and -1 if σ is odd, then the term corresponding to σ is

$$A_1^{\sigma(1)} A_2^{\sigma(2)} \dots A_n^{\sigma(n)} \,\mathrm{vol}(\boldsymbol{e}_{\sigma(1)}, \dots, \boldsymbol{e}_{\sigma(n)}) = (-1)^\sigma A_1^{\sigma(1)} A_2^{\sigma(2)} \dots A_n^{\sigma(n)}.$$

Summing over all permutations in the symmetric group S_n gives

$$\mathrm{vol}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n) = \sum_{\sigma \in S_n} (-1)^\sigma A_1^{\sigma(1)} A_2^{\sigma(2)} \dots A_n^{\sigma(n)}.$$

(Existence). It remains to show the preceding satisfies (i)–(iii).

Condition (i) follows from the distributive law for real numbers, because each summand contains precisely one factor coming from $\boldsymbol{v}_1$.

To prove Condition (ii), let $\tau = (i\ j)$ exchange i and j, and note that $(-1)^\sigma = -(-1)^{\sigma\tau}$ because τ is odd. The general summand after swapping $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ is

$$(-1)^\sigma A_1^{\sigma\tau(1)} A_2^{\sigma\tau(2)} \dots A_n^{\sigma\tau(n)} = -(-1)^{\sigma\tau} A_1^{\sigma\tau(1)} A_2^{\sigma\tau(2)} \dots A_n^{\sigma\tau(n)}.$$

Summing over σ in S_n amounts to summing over $\sigma\tau$.

To prove (iii), note that $\boldsymbol{e}_j = \sum_i \delta_j^i \boldsymbol{e}_i$, so

$$\mathrm{vol}(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n) = \sum_{\sigma \in S_n} (-1)^\sigma \delta_1^{\sigma(1)} \delta_2^{\sigma(2)} \dots \delta_n^{\sigma(n)}.$$

The only summand with a non-zero contribution comes from the identity permutation, and the contribution is 1. □

Example 3.48. The box spanned by $\big((2,-3),(1,-4)\big)$ has signed area

$$\det \begin{bmatrix} 2 & 1 \\ -3 & -4 \end{bmatrix} = (2)(-4) - (1)(-3) = -5.$$

That is, the box (parallelogram) has area 5, and the small angle from $(2,-3)$ to $(1,-4)$ is oppositely oriented to the small angle from $\boldsymbol{e}_1$ to $\boldsymbol{e}_2$ (because the sign is negative).

Definition 3.49. The *determinant* of an $n \times n$ matrix $A = [A^i_j]$ is

$$\det A = \sum_{\sigma \in S_n} (-1)^\sigma A_1^{\sigma(1)} A_2^{\sigma(2)} \dots A_n^{\sigma(n)}.$$

Theorem 3.50. *If A is an $n \times n$ matrix, then* $\det A = \det A^\mathsf{T}$.

Proof. Since $(-1)^{\sigma^{-1}} = (-1)^\sigma$, summing over σ^{-1} in S_n gives

$$\begin{aligned} \det A &= \sum_{\sigma^{-1} \in S_n} (-1)^\sigma A_1^{\sigma^{-1}(1)} \dots A_n^{\sigma^{-1}(n)} \\ &= \sum_{\sigma \in S_n} (-1)^\sigma A^1_{\sigma(1)} A^2_{\sigma(2)} \dots A^n_{\sigma(n)}. \end{aligned}$$

The last sum is $\det A^\mathsf{T}$, since the (i,j) entry of A^T is A^j_i. □

Theorem 3.51. *If $A = [A^i_j]$ is upper triangular or lower triangular then the determinant is the product of the diagonal entries:*

$$\det A = \prod_{i=1}^n A^i_i.$$

Proof. Assume first that A is upper triangular, namely that $A^i_j = 0$ for $j < i$. In the polynomial formula for $\det A$, a summand

$$(-1)^\sigma A_1^{\sigma(1)} A_2^{\sigma(2)} \dots A_n^{\sigma(n)}$$

is zero unless $\sigma(j) \le j$ for each j. But the only permutation satisfying these conditions is the identity: If σ is not the identity, there is a smallest index j with $\sigma(j) \ne j$, and since σ is a bijection of the set $\{1,2,\dots,n\}$, $\sigma(j) > j$. Consequently, if A is upper triangular, then

$$\det A = \sum_{\sigma \in S_n} (-1)^\sigma A_1^{\sigma(1)} A_2^{\sigma(2)} \dots A_n^{\sigma(n)} = A^1_1 A^2_2 \dots A^n_n.$$

If A is lower triangular, then A^T is upper triangular. □

Determinants and Row Operations

The determinant of a square matrix is the signed volume of the box spanned by its columns or (because of Theorem 3.50) by its rows. (This is geometrically non-obvious; the *boxes themselves* do not generally have the same shape.)

Properties (i) and (ii) for the determinant, regarded as a function of the *rows* of an $m \times m$ matrix, specify how the determinant changes under an elementary row operation:

Theorem 3.52. *If A is an $m \times m$ matrix, E an elementary matrix, and $A = EA'$, then $\det A = (\det E)(\det A')$.*

Proof. We check the claim explicitly for row operations of each type.

(Type I). Adding a multiple of one row to another does not change the determinant:

$$\begin{aligned}
&\operatorname{vol}(\dots, \boldsymbol{v}_i + c\boldsymbol{v}_j, \dots, \boldsymbol{v}_j, \dots) \\
&\qquad = \operatorname{vol}(\dots, \boldsymbol{v}_i, \dots, \boldsymbol{v}_j, \dots) + c\operatorname{vol}(\dots, \boldsymbol{v}_j, \dots, \boldsymbol{v}_j, \dots) \\
&\qquad = \operatorname{vol}(\dots, \boldsymbol{v}_i, \dots, \boldsymbol{v}_j, \dots).
\end{aligned}$$

But a Type I. elementary matrix is triangular, with 1s on the diagonal, so $\det E = 1$ in this case.

(Type II). Multiplying the ith row by c multiplies the determinant by c. But the Type II. elementary matrix for this operation is diagonal, with single c in the (i, i) entry and 1s elsewhere on the diagonal, so $\det E = c$ in this case.

(Type III). Exchanging two rows of A multiplies $\det A$ by -1. But the corresponding Type III. elementary matrix E is obtained from the identity by swapping two rows, so $\det E = -\det I_m = -1$ in this case. □

Remark 3.53. As a fringe benefit of the proof, the determinant of an elementary matrix is non-zero.

Theorem 3.52 (together with induction on the number of factors) shows that if E is a product of elementary matrices, then (i) $\det E \neq 0$; (ii) $\det(EC) = (\det E)(\det C)$ for every $m \times m$ matrix C.

Corollary 3.54. *A matrix A is invertible if and only if $\det A \neq 0$.*

Proof. Put A in reduced row-echelon form, i.e., factor $A = EA'$ with E a product of elementary matrices and A' a reduced row-echelon matrix, necessarily upper triangular.

By Theorem 1.49, A is invertible if and only if $A' = I_m$, if and only if A' does not contain a row of 0s, if and only if $\det A' \neq 0$. The corollary follows since $\det E \neq 0$ and $\det A = (\det E)(\det A')$. $\square$

Corollary 3.55. *If A and B are $m \times m$ matrices, then*

$$\det(AB) = (\det A)(\det B).$$

In particular, $\det(AB) = \det(BA)$.

Proof. Factor $A = EA'$ as in the preceding proof. If A is invertible, then $A' = I_m$, and $\det(AB) = \det(EB) = \det(E)(\det B) = \det(A)(\det B)$.

If A is not invertible, then $\det A = 0$ by the preceding corollary. Further, the reduced row-echelon matrix A' has a row of 0s, so $A'B$ has a row of zeros. The reduced row-echelon matrix of $A'B$ consequently has a row of zeros, so $\det(A'B) = 0$, and

$$\det(AB) = (\det E)(\det A'B) = 0 = (\det A)(\det B). \qquad \square$$

Corollary 3.56. *If A is invertible, then* $\det(A^{-1}) = 1/\det A$.

Proof. We have $I_m = AA^{-1}$, so $1 = \det I_m = (\det A)(\det A^{-1})$. $\square$

Corollary 3.57. *If A is an arbitrary $m \times m$ matrix and P is invertible, then* $\det(P^{-1}AP) = \det A$.

Proof. We have $\det(P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A$. $\square$

Example 3.58. Find the signed volume of the box spanned by the ordered triple $\big((1,1,1),(1,2,4),(1,3,9)\big)$.

The task is to find the determinant of the matrix whose columns are the given vectors. Although there is a six-term formula for a 3×3 determinant, we use row operations for simplicity (and illustration). We also use vertical bars, $\det A = |A|$, to save space:

$$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{vmatrix} \xrightarrow[R_3 - R_1]{R_2 - R_1,} \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 3 & 8 \end{vmatrix} \xrightarrow{R_3 - 3R_2} \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{vmatrix} = (1)(1)(2) = 2.$$

For practice, check the answer with the six-term formula.

Example 3.59. If $a < b < c$ are real numbers, then

$$\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} \xrightarrow{\substack{R_2 - R_1, \\ R_3 - R_1}} \begin{vmatrix} 1 & a & a^2 \\ 0 & b-a & (b-a)(b+a) \\ 0 & c-a & (c-a)(c+a) \end{vmatrix}$$

$$\xrightarrow{\substack{\frac{1}{b-a}R_2, \\ \frac{1}{c-a}R_3}} (b-a)(c-a) \begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & (b+a) \\ 0 & 1 & (c+a) \end{vmatrix}$$

$$\xrightarrow{R_3 - R_2} (b-a)(c-a) \begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & (b+a) \\ 0 & 0 & (c-b) \end{vmatrix} = (b-a)(c-a)(c-b).$$

Exercises

Exercise 3.1. Let $\boldsymbol{a}$ be a non-zero vector in $\mathbf{R}^n$, and let λ be a non-zero real number.

(a) Use the orthogonal projection formula to show that

$$\operatorname{proj}_{\lambda\boldsymbol{a}}(\boldsymbol{x}) = \operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x}.$$

(In words, orthogonal projection to $\boldsymbol{a}$ depends only on the direction of $\boldsymbol{a}$, not on its magnitude.)

(b) If $\boldsymbol{a}$ is a unit vector, how does the formula for $\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{x})$ simplify?

Exercise 3.2. Each part refers to

$$\boldsymbol{u}_1 = \tfrac{1}{\sqrt{3}}(1,1,1), \qquad \boldsymbol{u}_2 = \tfrac{1}{\sqrt{2}}(1,0,-1), \qquad \boldsymbol{u}_3 = \tfrac{1}{\sqrt{6}}(1,-2,1).$$

(a) Verify that $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthonormal set.

(b) Show by direct calculation that if $\boldsymbol{x} = (x^1, x^2, x^3)$ is arbitrary, then

$$\boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{u}_1 \rangle \cdot \boldsymbol{u}_1 + \langle \boldsymbol{x}, \boldsymbol{u}_2 \rangle \cdot \boldsymbol{u}_2 + \langle \boldsymbol{x}, \boldsymbol{u}_3 \rangle \cdot \boldsymbol{u}_3.$$

Exercise 3.3. In $\mathbf{R}^4$, suppose $\boldsymbol{u}_1 = \frac{1}{2}(1,1,1,1)$, $\boldsymbol{u}_2 = \frac{1}{2}(1,-1,-1,1)$, $\boldsymbol{u}_3 = \frac{1}{2}(1,-1,1,-1)$, $\boldsymbol{u}_4 = \frac{1}{2}(1,1,-1,-1)$.

(a) Verify that $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4\}$ is an orthonormal set. (You need to check six inner products, in addition to computing four magnitudes.)

(b) Calculate the four outer products $(\boldsymbol{u}_j \boldsymbol{u}_j^\mathsf{T})_{j=1}^4$.

(c) Show by direct calculation that if $\boldsymbol{x} = (x^1, x^2, x^3, x^4)$ is arbitrary,

$$\boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{u}_1 \rangle \cdot \boldsymbol{u}_1 + \langle \boldsymbol{x}, \boldsymbol{u}_2 \rangle \cdot \boldsymbol{u}_2 + \langle \boldsymbol{x}, \boldsymbol{u}_3 \rangle \cdot \boldsymbol{u}_3 + \langle \boldsymbol{x}, \boldsymbol{u}_4 \rangle \cdot \boldsymbol{u}_4.$$

Exercise 3.4. Each part refers to the vectors $\boldsymbol{v}_1 = (1, 1, 1, 1)$ and $\boldsymbol{v}_2 = (1, 2, 3, 4)$ in $\mathbf{R}^4$, and the plane W they span.

(i) Use Gram-Schmidt to find an orthonormal basis of W.

(ii) If $\boldsymbol{x} = (x^1, x^2, x^3, x^4)$ is arbitrary, decompose $\boldsymbol{x}$ in $W \oplus W^\perp$.

(iii) Find a system of two equations in four variables whose solution set is W. Similarly for $W^\perp$.

Exercise 3.5. Referring to Example 3.41:

(a) Form the matrix P whose columns are $(\boldsymbol{u}_j)_{j=0}^3$ and verify that $P^\mathsf{T} P = I_4$ and $PP^\mathsf{T} = I_4$.

(b) Calculate the projection matrix $\mathrm{proj}_W = \boldsymbol{u}_1 \boldsymbol{u}_1^\mathsf{T} + \boldsymbol{u}_2 \boldsymbol{u}_2^\mathsf{T} + \boldsymbol{u}_3 \boldsymbol{u}_3^\mathsf{T}$ directly, and verify the result is consistent with Example 3.23.

Exercise 3.6. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors in $\mathbf{R}^n$. The *parallelogram* they span is the plane figure with vertices $\mathbf{0}^n$, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{x}+\boldsymbol{y}$. The *diagonals* are the segments from $\mathbf{0}^n$ to $\boldsymbol{x} + \boldsymbol{y}$ and from $\boldsymbol{x}$ to $\boldsymbol{y}$.

(a) Prove the *parallelogram law*:

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 + \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2\big(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\big),$$

and give a geometric interpretation.

(b) Prove the *polarization identity*:

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 4 \langle \boldsymbol{x}, \boldsymbol{y} \rangle .$$

Conclude that the diagonals have equal magnitude if and only if the parallelogram is a rectangle.

(c) Prove that the diagonals of a parallelogram are orthogonal if and only if the parallelogram is a rhombus.

Exercise 3.7. Let P be an $n \times n$ matrix. Prove that P is orthogonal if and only if $\|P\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for all $\boldsymbol{x}$ in $\mathbf{R}^n$.
Suggestion: One direction is immediate from Theorem 3.33. For the other direction, use the polarization identity (preceding exercise).

Chapter 4

Linear Transformations

4.1 Linear Transformations

Definition 4.1. Let V and W be arbitrary vector spaces. A mapping $T : V \to W$ is said to be a *linear transformation* if T satisfies the *morphism condition*:

$$T(c\boldsymbol{x} + \boldsymbol{y}) = cT(\boldsymbol{x}) + T(\boldsymbol{y}) \quad \text{for all } \boldsymbol{x},\ \boldsymbol{y} \text{ in } V \text{ and all real } c.$$

Remark 4.2. Geometrically, $T(c\boldsymbol{x} + \boldsymbol{y}) = cT(\boldsymbol{x}) + T(\boldsymbol{y})$ means T maps the parallelogram in V with sides $c\boldsymbol{x}$ and $\boldsymbol{y}$ to the parallelogram in W with sides $cT(\boldsymbol{x})$ and $T(\boldsymbol{y})$, Figure 4.1. The mapping T is a linear transformation if and only if T has this effect on *every* parallelogram.

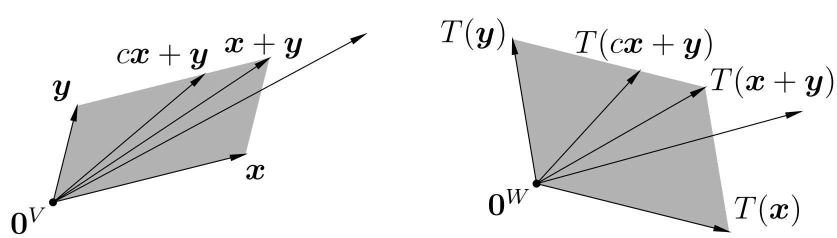


Figure 4.1: A linear transformation preserves parallelograms.

Remark 4.3. Induction on the number of summands proves that "a linear transformation distributes over linear combinations". Precisely, if $\boldsymbol{v}_1, \dots, \boldsymbol{v}_k$ are elements of V and $x^1, \dots, x^k$ are real numbers, then

$$T\left(\sum_{i=1}^{k} x^i \boldsymbol{v}_i\right) = \sum_{i=1}^{k} x^i T(\boldsymbol{v}_i).$$

AMS Open Math Notes: Works in Progress; Reference # OMN:201801.110759; Last Revised: 2018-01-20 09:27:31

Definition 4.4. Let V, V', and V'' be vector spaces, $T : V \to V'$ and $T' : V' \to V''$ linear transformations. The *composition* $T'T : V \to V''$ is defined by $T'T(\boldsymbol{x}) = T'\big(T(\boldsymbol{x})\big)$ for $\boldsymbol{x}$ in V.

Remark 4.5. A composition $T'T$ is linear, Exercise 4.8.

A linear transformation is the analog in linear algebra of a group homomorphism. Much of the terminology carries over, as do basic results about homomorphisms.

Definition 4.6. Let $T : V \to W$ be a linear transformation.

The *kernel* of T is the set

$$\ker(T) = \{\boldsymbol{x} \text{ in } V : T(\boldsymbol{x}) = \mathbf{0}^W\} \subseteq V.$$

The *image* of T is the set

$$\operatorname{im}(T) = T(V) = \{\boldsymbol{y} \text{ in } W : \boldsymbol{y} = T(\boldsymbol{x}) \text{ for some } \boldsymbol{x} \text{ in } V\} \subseteq W.$$

If, for all $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in V, $T(\boldsymbol{x}_1) = T(\boldsymbol{x}_2)$ implies $\boldsymbol{x}_1 = \boldsymbol{x}_2$, then T is said to be *injective*.

If $T(V) = W$, i.e., if, for every $\boldsymbol{y}$ in W, there exists an $\boldsymbol{x}$ in V such that $\boldsymbol{y} = T(\boldsymbol{x})$, then T is said to be *surjective*.

If T is bijective, i.e., both injective and surjective, then T is an *isomorphism* (of vector spaces).

Theorem 4.7. *Let $T : V \to W$ be a linear transformation.*

(i) *The kernel of T is a subspace of V.*

(ii) *The image of T is a subspace of W.*

(iii) *T is injective if and only if $\ker(T) = \{\mathbf{0}^V\}$, i.e., $\dim\ker(T) = 0$.*

(iv) *If T is an isomorphism, then the inverse map $T^{-1} : W \to V$ is an isomorphism.*

Proof. (i). We first show $T(\mathbf{0}^V) = \mathbf{0}^W$, proving the kernel is non-empty: If $\boldsymbol{x}$ is an arbitrary element of V, then

$$T(\mathbf{0}^V) = T(0 \cdot \boldsymbol{x}) = 0 \cdot T(\boldsymbol{x}) = \mathbf{0}^W.$$

If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are elements of $\ker(T)$ and c is real, then

$$T(c\boldsymbol{x}_1 + \boldsymbol{x}_2) = cT(\boldsymbol{x}_1) + T(\boldsymbol{x}_2) = c \cdot \mathbf{0}^W + \mathbf{0}^W = \mathbf{0}^W,$$

so $c\boldsymbol{x}_1 + \boldsymbol{x}_2 \in \ker(T)$. By Theorem 2.26, $\ker(T)$ is a subspace of V.

(ii). Suppose $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are elements of $T(V)$ and c is real. By hypothesis, there exist $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in V such that $T(\boldsymbol{x}_1) = \boldsymbol{y}_1$ and $T(\boldsymbol{x}_2) = \boldsymbol{y}_2$. Consequently,

$$T(c\boldsymbol{x}_1 + \boldsymbol{x}_2) = cT(\boldsymbol{x}_1) + T(\boldsymbol{x}_2) = c\boldsymbol{y}_1 + \boldsymbol{y}_2,$$

so $c\boldsymbol{y}_1 + \boldsymbol{y}_2 \in T(V)$. By Theorem 2.26, $T(V)$ is a subspace of W.

(iii). Suppose T is injective. If $\boldsymbol{x}$ is an arbitrary vector in $\ker(T)$, then $T(\boldsymbol{x}) = \mathbf{0}^W = T(\mathbf{0}^V)$, so $\boldsymbol{x} = \mathbf{0}^V$. That is, $\ker(T) = \{\mathbf{0}^V\}$.

Conversely, suppose $\ker(T) = \{\mathbf{0}^V\}$. If $T(\boldsymbol{x}_1) = T(\boldsymbol{x}_2)$, we have $\mathbf{0}^W = T(\boldsymbol{x}_1) - T(\boldsymbol{x}_2) = T(\boldsymbol{x}_1 - \boldsymbol{x}_2)$, i.e., $\boldsymbol{x}_1 - \boldsymbol{x}_2 \in \ker(T) = \{\mathbf{0}^V\}$. That is, $\boldsymbol{x}_1 = \boldsymbol{x}_2$. Since $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ were arbitrary, T is injective.

(iv). The mapping $T^{-1} : W \to V$ is defined by $T^{-1}(\boldsymbol{y}) = \boldsymbol{x}$ if and only if $T(\boldsymbol{x}) = \boldsymbol{y}$. Bijectivity of T^{-1} is obvious; it suffices to prove T^{-1} is linear.

Suppose $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are elements of W, c is real, and $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the unique vectors in V such that $T(\boldsymbol{x}_1) = \boldsymbol{y}_1$ and $T(\boldsymbol{x}_2) = \boldsymbol{y}_2$. Since

$$c\boldsymbol{y}_1 + \boldsymbol{y}_2 = cT(\boldsymbol{x}_1) + T(\boldsymbol{x}_2) = T(c\boldsymbol{x}_1 + \boldsymbol{x}_2)$$

by linearity of T, the definition of T^{-1} gives

$$T^{-1}(c\boldsymbol{y}_1 + \boldsymbol{y}_2) = c\boldsymbol{x}_1 + \boldsymbol{x}_2 = cT^{-1}(\boldsymbol{y}_1) + T^{-1}(\boldsymbol{y}_2). \qquad \square$$

Corollary 4.8. *If V is finite-dimensional and $T : V \to W$ is a linear transformation, then $\dim T(V) \leq \dim V$, with equality if and only if T is injective.*

Proof. If $(\boldsymbol{v}_j)_{j=1}^n$ is a basis of V, and if we put $\boldsymbol{w}_j = T(\boldsymbol{v}_j)$, then the n-element set $(\boldsymbol{w}_j)_{j=1}^n$ spans $T(V)$: Every $\boldsymbol{x}$ in V may be written uniquely as $\sum_j x^j \boldsymbol{v}_j$, and by linearity, $T(\boldsymbol{x}) = \sum_j x^j \boldsymbol{w}_j$. Theorem 2.64 implies $\dim T(V) \leq \dim V$.

Equality of dimensions holds if and only if $(\boldsymbol{w}_j)_{j=1}^n$ is linearly independent, if and only if no non-trivial linear combination $\sum_j x^j \boldsymbol{w}_j$ is $\mathbf{0}^W$, if and only if no non-trivial linear combination $\sum_j x^j \boldsymbol{v}_j$ is in $\ker(T)$, if and only if T is injective. $\square$

Remark 4.9. In words, applying a linear transformation T cannot increase dimension, and strictly reduces dimension unless T is injective.

Examples of Linear Transformations

Example 4.10. Let V and W be arbitrary vector spaces. The *zero map* $\mathbf{0}_V^W : V \to W$, defined by $\mathbf{0}_V^W(\boldsymbol{x}) = \mathbf{0}^W$ for all $\boldsymbol{x}$ in V, is linear.

Example 4.11. Let V be an arbitrary vector space. The *identity map* $I_V : V \to V$, defined by $I_V(\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x}$ in V, is linear.

Example 4.12. If $A = [A_j^i]$ is an $m \times n$ real matrix, multiplication by A defines a linear transformation $\mu_A : \mathbf{R}^n \to \mathbf{R}^m$ via $\mu_A(\boldsymbol{x}) = A\boldsymbol{x}$. In components, if $\boldsymbol{x} = [x^j]$, and $(\boldsymbol{e}_i)_{i=1}^m$ is the standard basis of $\mathbf{R}^m$, then

$$\mu_A(\boldsymbol{x}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_j^i x^j \right) \boldsymbol{e}_i.$$

Linearity follows from properties of matrix operations, Theorem 1.19:

$$\mu_A(c\boldsymbol{x} + \boldsymbol{y}) = A(c\boldsymbol{x} + \boldsymbol{y}) = c(A\boldsymbol{x}) + A\boldsymbol{y} = c\mu_A(\boldsymbol{x}) + \mu_A(\boldsymbol{y}).$$

Conversely, every linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ can be written as matrix multiplication in this way, Theorem 4.26.

Theorem 4.13. *Let $(V, +, \cdot)$ be an arbitrary finite-dimensional vector space of dimension n, and let $S = (\boldsymbol{v}_j)_{j=1}^n$ be a basis. The mapping $\iota^S : V \to \mathbf{R}^n$ that sends each $\boldsymbol{x}$ in V to its coordinate vector $\iota^S(\boldsymbol{x}) = [\boldsymbol{x}]^S$ in $\mathbf{R}^n$ is an isomorphism.*

Proof. (ι^S is linear). If $\boldsymbol{x}$ is an arbitrary element of V, then by definition, $[\boldsymbol{x}]^S = [x^j]$ if and only if

$$\boldsymbol{x} = x^1\boldsymbol{v}_1 + \cdots + x^n\boldsymbol{v}_n = \sum_{j=1}^{n} x^j \boldsymbol{v}_j.$$

If $\boldsymbol{y}$ is an arbitrary element of V with $[\boldsymbol{y}]^S = [y^j]$, and if c is real, then

$$c\boldsymbol{x} + \boldsymbol{y} = c\sum_{j=1}^{n} x^j \boldsymbol{v}_j + \sum_{j=1}^{n} y^j \boldsymbol{v}_j = \sum_{j=1}^{n} (cx^j + y^j)\boldsymbol{v}_j.$$

In terms of coordinate vectors,

$$[c\boldsymbol{x} + \boldsymbol{y}]^S = [cx^j + y^j] = c[x^j] + [y^j] = c[\boldsymbol{x}]^S + [\boldsymbol{y}]^S.$$

(ι^S is bijective). The inverse mapping $(\iota^S)^{-1} : \mathbf{R}^n \to V$ sends an ordered n-tuple $[x^j]$ to the linear combination from S with the corresponding coefficients,

$$(\iota^S)^{-1}\big([x^j]\big) = \sum_{j=1}^{n} x^j \boldsymbol{v}_j.$$

Existence of an inverse proves bijectivity. □

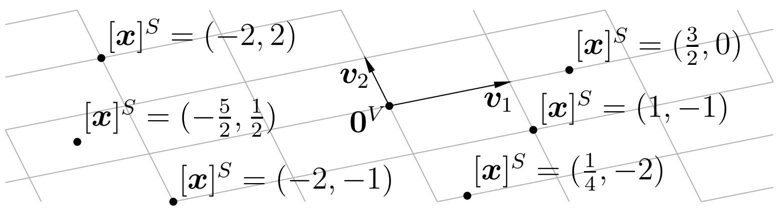


Figure 4.2: Coordinate vectors in a two-dimensional vector space.

Remark 4.14. Geometrically, a basis S with n elements defines a *coordinate grid* consisting of linear combinations where at least $(n-1)$ coefficients are integers, Figure 4.2; the isomorphism ι^S assigns to each location $\boldsymbol{x}$ in V an "address" $[\boldsymbol{x}]^S$ in $\mathbf{R}^n$.

Conceptually, the mapping ι^S and its inverse define a "dictionary" between V and $\mathbf{R}^n$ in the presence of a basis S of V. This dictionary depends on S. Material below on "change of basis" tells us how to "translate" between the dictionaries associated to two bases.

Example 4.15. Let $V = C^\infty(\mathbf{R}, \mathbf{R})$ be the vector space of smooth functions under ordinary addition and scalar multiplication. The following define linear transformations from V to V (for (i)–(iii)) or from V to $\mathbf{R}$ (for (iv), (v)). Linearity follows from theorems of calculus; a and b denote arbitrary real numbers.

(i) Differentiation, $D(f) = f'$.

(ii) Integration from a, $I_a(f) = \int_a^x f(t)\,dt$.

(iii) Multiplication by a smooth function ϕ, $\mu_\phi(f) = \phi f$.

(iv) Evaluation at a, $E_a(f) = f(a)$.

(v) Integration over $[a, b]$, $I_a^b(f) = \int_a^b f(t)\,dt$.

Linear Transformations on the Plane

Axial scaling, rotation, projection, and reflection are linear transformations $T : \mathbf{R}^2 \to \mathbf{R}^2$. If $(\boldsymbol{e}_j)_{j=1}^2$ is the standard basis and if $(\boldsymbol{w}_j)_{j=1}^2$ is an arbitrary pair of vectors (possibly equal), there exists a unique 2×2 matrix A satisfying $A\boldsymbol{e}_1 = \boldsymbol{w}_1$, and $A\boldsymbol{e}_2 = \boldsymbol{w}_2$, namely the matrix whose columns are $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. Figure 4.3 shows the geometric effect on the plane.

Example 4.16. (Axial scaling) Let a_1 and a_2 be real numbers, and define

$$T(x^1, x^2) = (a_1x^1, a_2x^2) = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}.$$

Here, $\boldsymbol{w}_i = a_i\boldsymbol{e}_i$. The geometric effect is to "scale" each coordinate axis. Axis-aligned rectangles map to axis-aligned rectangles under T, but if $|a_1| \neq |a_2|$, "tilted" rectangles do not map to rectangles.

Example 4.17. Let θ be real. Rotating the plane counterclockwise about the origin maps the standard basis vector $\boldsymbol{e}_1$ to $\boldsymbol{w}_1 = (\cos\theta, \sin\theta)$, and maps $\boldsymbol{e}_2$ to $\boldsymbol{w}_2 = (-\sin\theta, \cos\theta)$. Rotation preserves parallelograms, and is therefore a linear transformation Rot_θ, given by

$$\mathrm{Rot}_\theta(x^1, x^2) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = \begin{bmatrix} x^1\cos\theta - x^2\sin\theta \\ x^1\sin\theta + x^2\cos\theta \end{bmatrix}.$$

Example 4.18. Let θ be real, $\boldsymbol{a} = (\cos\theta, \sin\theta)$. Orthogonal projection $\mathrm{proj}_{\boldsymbol{a}}$ maps $\boldsymbol{e}_1$ to $\boldsymbol{w}_1 = \cos\theta(\cos\theta, \sin\theta)$, and maps $\boldsymbol{e}_2$ to $\boldsymbol{w}_2 = \sin\theta(\cos\theta, \sin\theta)$. The matrix of this linear transformation is

$$\begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \end{bmatrix}.$$

The factorization on the right is an *outer product*.

Example 4.19. Let θ be real, $\boldsymbol{a} = (\cos\theta, \sin\theta)$, $\boldsymbol{a}^\perp = (-\sin\theta, \cos\theta)$. Reflection across $\boldsymbol{a}$ is defined by

$$\mathrm{Ref}_\theta(\boldsymbol{x}) = \boldsymbol{x} - 2\,\mathrm{proj}_{\boldsymbol{a}^\perp}(\boldsymbol{x}) = -\boldsymbol{x} + 2\,\mathrm{proj}_{\boldsymbol{a}}(\boldsymbol{x}).$$

The matrix of this linear transformation may be found by calculating $\mathrm{Ref}_\theta(\boldsymbol{e}_1)$ and $\mathrm{Ref}_\theta(\boldsymbol{e}_2)$, or by using the preceding example:

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} + 2\begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix} = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}.$$

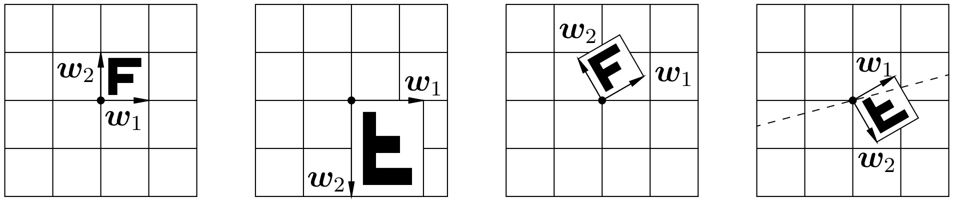


Figure 4.3: The identity, axial scaling, rotation, and reflection.

4.2 The Space of Linear Transformations

Proposition 4.20. *If T_1 and $T_2 : V \to W$ are linear transformations and α is real, the linear combination $\alpha T_1 + T_2 : V \to W$ defined by*

$$(\alpha T_1 + T_2)(\boldsymbol{x}) = \alpha T_1(\boldsymbol{x}) + T_2(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \text{ in } V$$

is a linear transformation.

Proof. If $\boldsymbol{x}$ and $\boldsymbol{y}$ are arbitrary elements of V and c is real, then

$$\begin{aligned}(\alpha T_1 + T_2)(c\boldsymbol{x} + \boldsymbol{y}) &= \alpha T_1(c\boldsymbol{x} + \boldsymbol{y}) + T_2(c\boldsymbol{x} + \boldsymbol{y}) \\ &= \alpha\big[cT_1(\boldsymbol{x}) + T_1(\boldsymbol{y})\big] + \big[cT_2(\boldsymbol{x}) + T_2(\boldsymbol{y})\big] \\ &= c\big[\alpha T_1(\boldsymbol{x}) + T_2(\boldsymbol{x})\big] + \big[\alpha T_1(\boldsymbol{y}) + T_2(\boldsymbol{y})\big] \\ &= c(\alpha T_1 + T_2)(\boldsymbol{x}) + (\alpha T_1 + T_2)(\boldsymbol{y}).\end{aligned}$$ □

Remark 4.21. Let V and W be vector spaces. The set $\mathcal{L}(V, W)$ of linear transformations $T : V \to W$ becomes a vector space under the operations of addition and scalar multiplication defined in the proposition.

If $V = \mathbf{R}^n$ and $W = \mathbf{R}^m$, it turns out that $\mathcal{L}(V, W) = \mathbf{R}^{m\times n}$.

Proposition 4.22. *Let V be a vector space with basis $S = (\boldsymbol{v}_\alpha)$. If $T : V \to W$ is a linear transformation satisfying $T(\boldsymbol{v}_\alpha) = \mathbf{0}^W$ for all α, then $T(\boldsymbol{x}) = \mathbf{0}^W$ for all $\boldsymbol{x}$ in V.*

The notation $(\boldsymbol{v}_\alpha)$ indicates an arbitrary family of vectors; that is, the proposition holds even for infinite-dimensional spaces.

Proof. Let $\boldsymbol{x}$ be an arbitrary element of V, and write $\boldsymbol{x}$ as a linear combination of basis elements. Since a linear combination is finite by definition, we have

$$\boldsymbol{x} = \sum_{j=1}^{n} x^j \boldsymbol{v}_j$$

for some basis vectors $\boldsymbol{v}_j$ in S and some scalars x^j. By linearity,

$$T(\boldsymbol{x}) = T\left(\sum_{j=1}^{n} x^j \boldsymbol{v}_j\right) = \sum_{j=1}^{n} x^j T(\boldsymbol{v}_j) = \sum_{j=1}^{n} x^j \cdot \mathbf{0}^W = \mathbf{0}^W. \qquad \square$$

Corollary 4.23. *If S is a basis of $(V, +, \cdot)$, and if T_1 and T_2 are linear transformations from V to W that agree on S, in that $T_1(\boldsymbol{v}_\alpha) = T_2(\boldsymbol{v}_\alpha)$ for all $\boldsymbol{v}_\alpha$ in S, then $T_1(\boldsymbol{x}) = T_2(\boldsymbol{x})$ for all $\boldsymbol{x}$ in V.*

Proof. Apply Proposition 4.22 to $T_1 - T_2$. $\square$

Theorem 4.24. *Let V be an n-dimensional vector space, $S = (\boldsymbol{v}_j)_{j=1}^n$ a basis, W an arbitrary vector space, and $S' = (\boldsymbol{w}_j)_{j=1}^n$ an arbitrary ordered set of vectors in W.*

(i) *There exists a unique linear transformation $T : V \to W$ satisfying $T(\boldsymbol{v}_j) = \boldsymbol{w}_j$ for $j = 1, \dots, n$.*

(ii) *T is injective if and only if S' is linearly independent.*

(iii) *T is surjective if and only if S' spans W.*

Remark 4.25. In words, (i) says a linear transformation is uniquely specified by its values on a basis. If ordered sets S and S' of vectors are given as in the theorem, the transformation T is said to be obtained from the conditions $T(\boldsymbol{v}_j) = \boldsymbol{w}_j$ via *extension by linearity.*

Proof. (i). Because S is a basis of V, every vector $\boldsymbol{x}$ in V may be written uniquely as a linear combination $\sum_j x^j \boldsymbol{v}_j$ from S. If T is linear, then

$$T(\boldsymbol{x}) = \sum_{j=1}^{n} x^j T(\boldsymbol{v}_j) = \sum_{j=1}^{n} x^j \boldsymbol{w}_j.$$

This formula defines a mapping $T : V \to W$, easily checked to be linear, and to satisfy $T(\boldsymbol{v}_j) = \boldsymbol{w}_j$.

Uniqueness is asserted by Corollary 4.23.

In the remainder of the proof, the preceding formula, which expresses $T(\boldsymbol{x})$ as the general linear combination from S', is used freely.

(ii). A vector $\boldsymbol{x}$ is in $\ker(T)$ if and only if $\mathbf{0}^W = T(\boldsymbol{x}) = \sum x^j \boldsymbol{w}_j$. But T is injective if and only if $\ker(T) = \{\mathbf{0}^V\}$, if and only if the equation $\mathbf{0}^W = \sum x^j \boldsymbol{w}_j$ has only the trivial solution $[x^j] = \mathbf{0}^n$, if and only if S' is linearly independent.

(iii). A vector $\boldsymbol{y}$ is in $T(V)$ if and only if there exist scalars x^j such that $\boldsymbol{y} = T(\boldsymbol{x}) = \sum x^j \boldsymbol{w}_j$. But T is surjective if and only if $T(V) = W$, if and only if the equation $\boldsymbol{y} = \sum x^j \boldsymbol{w}_j$ has a solution $[x^j]$ for every $\boldsymbol{y}$ in W, if and only if S' spans W. □

The Matrix of a Transformation

We now have enough machinery to express an arbitrary linear transformation between finite-dimensional vector spaces in terms of matrices of real numbers.

Theorem 4.26. *Let V and W be finite-dimensional vector spaces, and assume $S = (\boldsymbol{v}_j)_{j=1}^n$ and $S' = (\boldsymbol{w}_i)_{i=1}^m$ are bases for V and W, respectively. If $T : V \to W$ is a linear transformation, there exist unique scalars A_j^i, for $i = 1, \dots, m$ and $j = 1, \dots, n$, such that*

$$T(\boldsymbol{v}_j) = \sum_{i=1}^{m} A_j^i \boldsymbol{w}_i \quad \textit{for } j = 1, \dots, n.$$

Proof. For each j, the vector $T(\boldsymbol{v}_j)$ can be written uniquely as a linear combination from S'. If A_j^i is the $\boldsymbol{w}_i$-component of $T(\boldsymbol{v}_j)$, then the equation in the theorem holds. □

Definition 4.27. The $m \times n$ matrix $[T]_S^{S'} = [A_j^i]$ is called the *matrix of T with respect to S and S'*.

Remark 4.28. If bases S for V and S' for W are fixed, this association of a matrix to a linear transformation defines a linear isomorphism $\iota_S^{S'}$ from the vector space $\mathcal{L}(V, W)$ to the vector space $\mathbf{R}^{m \times n}$ of $m \times n$ real matrices. Theorem 4.26 is the analog, for linear transformations, of the "coordinatization" in Theorem 4.13 for vectors.

Example 4.29. Let V be a vector space with basis $S = (\boldsymbol{v}_j)_{j=1}^n$. The matrix of the identity transformation is

$$[I_V]_S^S = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

whose jth column encodes the equation $I_V(\boldsymbol{v}_j) = \boldsymbol{v}_j$.

Example 4.30. Let V and W be finite-dimensional vector spaces with respective bases $S = (\boldsymbol{v}_j)_{j=1}^n$ and $S' = (\boldsymbol{w}_i)_{i=1}^m$. For each pair (i, j) with $i = 1, \dots, m$ and $j = 1, \dots, n$, let $E_i^j : V \to W$ be the linear transformation defined by $E_i^j(\boldsymbol{v}_j) = \boldsymbol{w}_i$ and $E_i^j(\boldsymbol{v}_k) = \mathbf{0}^W$ if $k \neq j$; that is, $E_i^j(\boldsymbol{v}_k) = \delta_k^j \boldsymbol{w}_i$. The matrix $[E_i^j]_S^{S'}$ is $\boldsymbol{e}_i^j = \boldsymbol{e}_i \boldsymbol{e}^j$, having a 1 in the (i, j) entry and 0's elsewhere.

Example 4.31. If $T : \mathbf{R}^n \to \mathbf{R}^m$ has one-dimensional image, there exist non-zero vectors $\boldsymbol{v}$ in $\mathbf{R}^n$ and $\boldsymbol{w}$ in $\mathbf{R}^m$ such that

$$T(\boldsymbol{x}) = \langle \boldsymbol{v}, \boldsymbol{x} \rangle \cdot \boldsymbol{w} = \boldsymbol{w}\boldsymbol{v}^{\mathsf{T}}\boldsymbol{x} \quad \text{for all } \boldsymbol{x} \text{ in } \mathbf{R}^n.$$

The standard matrix of T is $\boldsymbol{w}\boldsymbol{v}^{\mathsf{T}}$. The kernel and image of T are $\ker(T) = \operatorname{Span}(\boldsymbol{v})^{\perp}$ and $\operatorname{im}(T) = \operatorname{Span}(\boldsymbol{w})$.

Proof. By hypothesis, $T(V)$ is 1-dimensional. Let $\boldsymbol{w}$ be an arbitrary non-zero vector in the image. If $(\boldsymbol{e}_j)_{j=1}^n$ is the standard basis of $\mathbf{R}^n$, then for each j there is a scalar v^j such that $T(\boldsymbol{e}_j) = v^j \boldsymbol{w}$. Put $\boldsymbol{v} = [v^j]$.

If $\boldsymbol{x} = \sum_j x^j \boldsymbol{e}_j$ is an arbitrary vector in $\mathbf{R}^n$, then

$$T(\boldsymbol{x}) = \sum_{j=1}^n x^j T(\boldsymbol{e}_j) = \sum_{j=1}^n x^j v^j \boldsymbol{w} = \langle \boldsymbol{v}, \boldsymbol{x} \rangle \cdot \boldsymbol{w}. \qquad \square$$

Corollary 4.32. *In the notation of the theorem,*

$$[T(\boldsymbol{x})]^{S'} = [T]_S^{S'} [\boldsymbol{x}]^S \quad \textit{for all } \boldsymbol{x} \textit{ in } V.$$

Proof. Write $[T]_S^{S'} = [A_j^i]$ and $[\boldsymbol{x}]^S = [x^j]$, so that $\boldsymbol{x} = \sum_j x^j \boldsymbol{v}_j$. We want to express $T(\boldsymbol{x})$ with respect to S'. Substituting,

$$T(\boldsymbol{x}) = \sum_{j=1}^n x^j T(\boldsymbol{v}_j) = \sum_{j=1}^n x^j \sum_{i=1}^m A_j^i \boldsymbol{w}_i = \sum_{i=1}^m \left(\sum_{j=1}^n A_j^i x^j \right) \boldsymbol{w}_i.$$

The expression in parentheses is precisely the ith row of the matrix product $[T]_S^{S'} [\boldsymbol{x}]^S$, and is also the $\boldsymbol{w}_i$-component of $T(\boldsymbol{x})$, i.e., the ith row of $[T(\boldsymbol{x})]^{S'}$. $\square$

Corollary 4.33. *If $T : V \to V'$ and $T' : V' \to V''$ are linear transformations between finite-dimensional vector spaces with respective bases S, S', and S'', then*

$$[T'T]_S^{S''} = [T']_{S'}^{S''} [T]_S^{S'}.$$

Proof. For all $\boldsymbol{x}$ in V, Corollary 4.32 gives

$$[T'T]_S^{S''}[\boldsymbol{x}]^S = [T'T(\boldsymbol{x})]^{S''} = [T']_{S'}^{S''}[T(\boldsymbol{x})]_S^{S'} = [T']_{S'}^{S''}[T]_S^{S'}[\boldsymbol{x}]^S.$$

By Corollary 1.23, $[T'T]_S^{S''} = [T']_{S'}^{S''}[T]_S^{S'}$. □

Remark 4.34. In words, evaluation of a linear transformation and composition of transformations correspond to matrix multiplication. These results are a basic computational idiom in linear algebra, and justify the definition of matrix multiplication.

The conclusions of Corollaries 4.32 and 4.33 may be expressed, respectively, as *commutative diagrams*, collections of spaces and mappings in which two compositions with the same domain and target are the same mapping. Specifically:

$$\begin{tikzcd} \boldsymbol{x} \in V \arrow[r, "T"] \arrow[d, "\iota^S"'] & T(\boldsymbol{x}) \in W \arrow[d, "\iota^{S'}"] \\ {[\boldsymbol{x}]^S \in \mathbf{R}^n} \arrow[r, "{[T]_S^{S'}}"] & {[T(\boldsymbol{x})]^{S'} \in \mathbf{R}^m} \end{tikzcd} \qquad \begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\iota^S"] & V' \arrow[r, "T'"] \arrow[d, "\iota^{S'}"] & V'' \arrow[d, "\iota^{S''}"] \\ \mathbf{R}^n \arrow[r, "{[T]_S^{S'}}"] & \mathbf{R}^{n'} \arrow[r, "{[T']_{S'}^{S''}}"] & \mathbf{R}^{n''} \end{tikzcd}$$

Change of Basis

Definition 4.35. Let V be an n-dimensional vector space with bases S and S'. The $n \times n$ matrix $[I_V]_S^{S'}$ is called the *transition matrix from S to S'*.

Remark 4.36. By Corollary 4.32, $[\boldsymbol{x}]^{S'} = [I_V]_S^{S'}[\boldsymbol{x}]^S$ for all $\boldsymbol{x}$ in V. That is, the transition matrix "converts" coordinate vectors with respect to S into coordinate vectors with respect to S'.

It may help to think of an element $\boldsymbol{x}$ in V as an object, and to think of the coordinate vector $[\boldsymbol{x}]^S$ as a description of the object in the language S. The transition matrix $[I_V]_S^{S'}$ is a "dictionary" or "translator" from language S to language S'.

Corollary 4.37. *Let V be an n-dimensional vector space with bases S and S'.*

(i) *The transition matrix $[I_V]_S^{S'}$ is invertible, and the inverse is the transition matrix $[I_V]_{S'}^S$.*

(ii) *If $T : V \to V$ is linear, then $[T]_{S'}^{S'} = [I_V]_S^{S'}[T]_S^S[I_V]_{S'}^S$.*

Proof. Example 4.29 notes that the matrix of the identity transformation with respect to a *single* basis is the $n \times n$ identity matrix I_n. By Corollary 4.33,

$$[I_V]_{S'}^{S}[I_V]_{S}^{S'} = [I_V]_{S}^{S} = I_n, \qquad [I_V]_{S}^{S'}[I_V]_{S'}^{S} = [I_V]_{S'}^{S'} = I_n.$$

Assertion (ii) follows immediately from Corollary 4.33. □

Definition 4.38. Let n be a positive integer, and let A_1 and A_2 be $n \times n$ matrices. We say A_2 is *similar* to A_1 if there exists an invertible $n \times n$ matrix P such that $A_2 = P^{-1}A_1P$.

Remark 4.39. Corollary 4.37 (ii) says that if T is a linear operator on V (a linear transformation from V to itself), then the matrices of T with respect to arbitrary bases are similar.

Proposition 4.40. *Similarity is an equivalence relation on the set of $n \times n$ matrices.*

Proof. (Reflexivity). Since the identity matrix I_n is its own inverse, $A = I_nAI_n = I_n^{-1}AI_n$.

(Symmetry). Suppose $A_2 = P^{-1}A_1P$. Multiplying on the left by $(P^{-1})^{-1}$ and on the right by P^{-1} gives $(P^{-1})^{-1}A_2P^{-1} = A_1$. In words, if A_2 is similar to A_1, then A_1 is similar to A_2.

(Transitivity). Suppose $A_2 = P^{-1}A_1P$ and $A_3 = Q^{-1}A_2Q$. Substituting the first into the second,

$$A_3 = Q^{-1}A_2Q = Q^{-1}P^{-1}A_1PQ = (PQ)^{-1}A_1(PQ),$$

so A_3 is similar to A_1. □

Remark 4.41. Detecting whether two $n \times n$ matrices are similar or not requires work. Chapter 5 is devoted to locating, with the similarity class of a matrix A, a "proxy" having a particularly simple structure. Two matrices are similar if and only if they have the same proxy.

Example 4.42. Let $V = P_3$ be the four-dimensional vector space of cubic polynomials, $S = (1, t, t^2, t^3) = (\boldsymbol{v}_j)_{j=1}^4$ the "standard" basis, $S' = (1, 1+t, 1+t+t^2, 1+t+t^2+t^3) = (\boldsymbol{v}_j')_{j=1}^4$ a non-standard basis, and $D : P_3 \to P_3$ the derivative operator from calculus, $D(p) = p'$. We will find the transition matrices between S and S', calculate the matrix

of D with respect to S, and use Corollary 4.37 to find the matrix of D with respect to S'.

The jth column of the transition matrix $[I]^S_{S'}$ is the coordinate vector $[\boldsymbol{v}'_j]^S$ of the jth element of S' with respect to S. By inspection,

$$[I]^S_{S'} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The inverse can be computed either by calculating the coordinate vectors $[\boldsymbol{v}_j]^{S'}$, or by using row operations on the preceding matrix. Both are shown for illustration.

To calculate coordinate vectors, set up the vector equations

$$\begin{aligned}
\boldsymbol{v}_1 &= \underline{\qquad}\,\boldsymbol{v}'_1 + \underline{\qquad}\,\boldsymbol{v}'_2 + \underline{\qquad}\,\boldsymbol{v}'_3 + \underline{\qquad}\,\boldsymbol{v}'_4, \\
\boldsymbol{v}_2 &= \underline{\qquad}\,\boldsymbol{v}'_1 + \underline{\qquad}\,\boldsymbol{v}'_2 + \underline{\qquad}\,\boldsymbol{v}'_3 + \underline{\qquad}\,\boldsymbol{v}'_4, \\
\boldsymbol{v}_3 &= \underline{\qquad}\,\boldsymbol{v}'_1 + \underline{\qquad}\,\boldsymbol{v}'_2 + \underline{\qquad}\,\boldsymbol{v}'_3 + \underline{\qquad}\,\boldsymbol{v}'_4, \\
\boldsymbol{v}_4 &= \underline{\qquad}\,\boldsymbol{v}'_1 + \underline{\qquad}\,\boldsymbol{v}'_2 + \underline{\qquad}\,\boldsymbol{v}'_3 + \underline{\qquad}\,\boldsymbol{v}'_4.
\end{aligned}$$

In general, each equation becomes a system; here, the system is easy to solve by inspection:

$$\begin{aligned}
1 &= \underline{\ \ 1\ \ }\,\boldsymbol{v}'_1 + \underline{\ \ 0\ \ }\,\boldsymbol{v}'_2 + \underline{\ \ 0\ \ }\,\boldsymbol{v}'_3 + \underline{\ \ 0\ \ }\,\boldsymbol{v}'_4, \\
t &= \underline{\ -1\ }\,\boldsymbol{v}'_1 + \underline{\ \ 1\ \ }\,\boldsymbol{v}'_2 + \underline{\ \ 0\ \ }\,\boldsymbol{v}'_3 + \underline{\ \ 0\ \ }\,\boldsymbol{v}'_4, \\
t^2 &= \underline{\ \ 0\ \ }\,\boldsymbol{v}'_1 + \underline{\ -1\ }\,\boldsymbol{v}'_2 + \underline{\ \ 1\ \ }\,\boldsymbol{v}'_3 + \underline{\ \ 0\ \ }\,\boldsymbol{v}'_4, \\
t^3 &= \underline{\ \ 0\ \ }\,\boldsymbol{v}'_1 + \underline{\ \ 0\ \ }\,\boldsymbol{v}'_2 + \underline{\ -1\ }\,\boldsymbol{v}'_3 + \underline{\ \ 1\ \ }\,\boldsymbol{v}'_4.
\end{aligned}$$

The coefficients in the jth *row* are the jth column of $[I]^{S'}_S$.

Alternatively, use the matrix inversion algorithm:

$$\left[\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\substack{R_1-R_2,\\ R_2-R_3,\\ R_3-R_4}} \left[\begin{array}{cccc|rrrr} 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right].$$

The right-hand block is $[I]^{S'}_S$.

To calculate the standard matrix of D, evaluate $D(\boldsymbol{v}_j)$, then express the result as a coordinate vector with respect to S. Here, $D(1) = 0$,

$D(t) = 1$, $D(t^2) = 2t$, and $D(t^3) = 3t^2$. The standard coordinate vectors may be read off by inspection, giving

$$[D]_S^S = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Finally, $[I]_S^{S'}[D]_S^S[I]_S^{S'} = [D]_{S'}^{S'}$, or

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 & -1 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The third column signifies that $D(\boldsymbol{v}_3') = -\boldsymbol{v}_1' + 2\boldsymbol{v}_2'$, while the fourth column signifies that $D(\boldsymbol{v}_4') = -\boldsymbol{v}_1' - \boldsymbol{v}_2' + 3\boldsymbol{v}_3'$.

4.3 The Rank-Nullity Theorem

Let V be an n-dimensional vector space. The qualitative structure of a linear transformation $T : V \to W$ is simple: "Some of the dimensions of V are sent to $\mathbf{0}^W$; the rest and mapped injectively". We prove two precise statements of this result, one very concrete, one abstract. A third proof strategy, intermediate in abstraction, may be found in Exercise 4.15.

Definition 4.43. Let $A = [A_j^i]$ be an $m \times n$ matrix, with rows $(A^i)_{i=1}^m$ in $(\mathbf{R}^n)^*$ and columns $(A_j)_{j=1}^n$ in $\mathbf{R}^m$.

The *nullspace* of A, $\mathrm{Null}(A)$, is the solution space of the homogeneous system $A\boldsymbol{x} = \mathbf{0}^m$. The *nullity* of A is the dimension of the nullspace of A.

The *column space* of A, $\mathrm{Col}(A)$, is $\mathrm{Span}(A_j)_{j=1}^n \subseteq \mathbf{R}^m$. The *column rank* of A is the dimension of the column space of A.

The *row space* of A, $\mathrm{Row}(A)$, is $\mathrm{Span}(A^i)_{i=1}^m \subseteq (\mathbf{R}^n)^*$. The *row rank* of A is the dimension of the row space of A.

Remark 4.44. If $\mu_A : \mathbf{R}^n \to \mathbf{R}^m$ is multiplication by A, then the nullspace of A is the kernel of μ_A, and the column space of A is the image of μ_A.

Lemma 4.45. *If A is an $m \times n$ matrix, E is a product of $m \times m$ elementary row matrices, and $A' = EA$, then*

(i) $\mathrm{Null}(A) = \mathrm{Null}(A')$.

(ii) $\dim \mathrm{Col}(A) = \dim \mathrm{Col}(A')$.

(iii) $\mathrm{Row}(A) = \mathrm{Row}(A')$.

Proof. (i). This restates Proposition 1.38. (Elementary row operations do not change the homogeneous solution space.)

(ii). Multiplication by E defines an invertible (hence injective) linear transformation $\mu_E : \mathbf{R}^m \to \mathbf{R}^m$ sending $\mathrm{Col}(A)$ to $\mathrm{Col}(A')$. Corollary 4.8 implies $\dim \mathrm{Col}(A) = \dim \mathrm{Col}(A')$.

(iii). Each elementary row operation replaces a set of rows with a set of linear combinations of the rows, so $\mathrm{Row}(A') \subseteq \mathrm{Row}(A)$. Since E^{-1} is a product of elementary matrices, reversing the roles of A and A' gives $\mathrm{Row}(A) \subseteq \mathrm{Row}(A')$. □

Theorem 4.46 (Rank-Nullity for Matrices)**.** *If A is an $m \times n$ matrix,*

(i) $\dim \mathrm{Col}(A) = \dim \mathrm{Row}(A)$.

(ii) $\dim \mathrm{Null}(A) + \dim \mathrm{Col}(A) = n$.

Proof. (i). Let A' be the reduced row-echelon form of A, and let r denote the number of non-zero rows of A', i.e., the number of leading 1's. Since the rows of A' are linearly independent, $\dim \mathrm{Row}(A') = r$.

Since every column of A' has at most r non-zero components, we have $\dim \mathrm{Col}(A') \leq r$. But the r columns containing leading 1's are obviously linearly independent, so $\dim \mathrm{Col}(A') = r$.

By Lemma 4.45,

$$\dim \mathrm{Col}(A) = \dim \mathrm{Col}(A') = r = \dim \mathrm{Row}(A') = \dim \mathrm{Row}(A).$$

(ii). By Theorem 2.76, $\dim \mathrm{Null}(A)$ is the number of free variables, namely $(n - r)$, the number of columns not containing a leading 1, while $\dim \mathrm{Col}(A) = r$ is the number of basic variables. The sum of these dimensions is n. □

Definition 4.47. If A is an $m \times n$ matrix, the *rank* of A is the common value $\dim \mathrm{Col}(A) = \dim \mathrm{Row}(A)$.

Proposition 4.48. *Let A be an $m \times n$ matrix. If the basic variables of the homogeneous system $A\boldsymbol{x} = \mathbf{0}^m$ are in columns $(k_j)_{j=1}^r$, the corresponding columns $(A_{k_j})_{j=1}^r$ are a basis of* $\mathrm{Col}(A)$.

Remark 4.49. That is, one way to find a basis of $\mathrm{Col}(A)$ is to row-reduce A, then select the columns *of A itself* corresponding to the leading 1's of the row-echelon matrix.

Proof. Let B be the $m \times r$ matrix whose jth column is A_{k_j}. (In words, discard all the "free variable" columns of A.) The reduced row-echelon form B' is I_r followed by $m - r$ rows of 0's, so

$$\dim \mathrm{Col}(A) = r = \dim \mathrm{Col}(B') = \dim \mathrm{Col}(B).$$

Since B has r columns, its columns are linearly independent, hence form a basis of $\mathrm{Col}(A)$. □

Quotient Vector Spaces

Our second proof of the Rank-Nullity Theorem is the direct analog of the First Homomorphism Theorem from group theory.

Definition 4.50. The *nullity* of T is $\dim \ker(T)$. The *rank* is $\dim T(V)$.

Proposition 4.51. *Let $(V, +, \cdot)$ be a vector space, K a subspace. The binary relation $\boldsymbol{x}_1 \equiv \boldsymbol{x}_2 \pmod{K}$ if and only if $\boldsymbol{x}_2 - \boldsymbol{x}_1 \in K$ is an equivalence relation on V.*

Proof. (Reflexivity). If $\boldsymbol{x} \in V$, then $\boldsymbol{x} - \boldsymbol{x} = \mathbf{0}^V \in K$; by definition, $\boldsymbol{x} \equiv \boldsymbol{x} \pmod{K}$.

(Symmetry). If $\boldsymbol{x}_1 \equiv \boldsymbol{x}_2 \pmod{K}$, then $\boldsymbol{x}_2 - \boldsymbol{x}_1 \in K$. Since K is closed under inversion, $\boldsymbol{x}_1 - \boldsymbol{x}_2 = -(\boldsymbol{x}_2 - \boldsymbol{x}_1) \in K$, so $\boldsymbol{x}_2 \equiv \boldsymbol{x}_1 \pmod{K}$.

(Transitivity). If $\boldsymbol{x}_1 \equiv \boldsymbol{x}_2 \pmod{K}$ and $\boldsymbol{x}_2 \equiv \boldsymbol{x}_3 \pmod{K}$, then by hypothesis, $\boldsymbol{x}_2 - \boldsymbol{x}_1 \in K$ and $\boldsymbol{x}_3 - \boldsymbol{x}_2 \in K$. Since K is closed under addition, $\boldsymbol{x}_3 - \boldsymbol{x}_1 = (\boldsymbol{x}_3 - \boldsymbol{x}_2) + (\boldsymbol{x}_2 - \boldsymbol{x}_1) \in K$, proving that $\boldsymbol{x}_1 \equiv \boldsymbol{x}_3 \pmod{K}$. □

Definition 4.52. Let $(V, +, \cdot)$ be a vector space, K a subspace. For each $\boldsymbol{x}_0$ in V, the set

$$\boldsymbol{x}_0 + K = \{\boldsymbol{x} \text{ in } V : \boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{v} \text{ for some } \boldsymbol{v} \text{ in } K\}$$

is called a *coset* of K in V.

Remark 4.53. The equivalence classes of the relation in Proposition 4.51 are precisely the cosets of K in V: If $\boldsymbol{x}$ and $\boldsymbol{x}_0$ are arbitrary vectors in V, then $\boldsymbol{x}_0 \equiv \boldsymbol{x} \pmod{K}$ if and only if $\boldsymbol{x} - \boldsymbol{x}_0 \in K$, if and only if $\boldsymbol{x} \in \boldsymbol{x}_0 + K$.

Lemma 4.54. *Let $(V, +, \cdot)$ be a vector space, K a subspace. If $\boldsymbol{x}_1 \equiv \boldsymbol{x}_2 \pmod{K}$ and $\boldsymbol{y}_1 \equiv \boldsymbol{y}_2 \pmod{K}$, and if c is real, then*

$$c\boldsymbol{x}_1 + \boldsymbol{y}_1 \equiv c\boldsymbol{x}_2 + \boldsymbol{y}_2 \pmod{K}.$$

Proof. By hypothesis, there exist elements $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ of K such that $\boldsymbol{x}_2 - \boldsymbol{x}_1 = \boldsymbol{x}_0$ and $\boldsymbol{y}_2 - \boldsymbol{y}_1 = \boldsymbol{y}_0$. Thus

$$(c\boldsymbol{x}_2 + \boldsymbol{y}_2) - (c\boldsymbol{x}_1 + \boldsymbol{y}_1) = c(\boldsymbol{x}_2 - \boldsymbol{x}_1) + (\boldsymbol{y}_2 - \boldsymbol{y}_1) = c\boldsymbol{x}_0 + \boldsymbol{y}_0 \in K. \quad \square$$

Remark 4.55. Lemma 4.54 guarantees that addition and scalar multiplication of *cosets* is “well-defined”, i.e., depends only on the cosets, and not on the representative elements of V.

The coset $K = \mathbf{0}^V + K$ acts as identity element for addition, i.e., as the zero vector. The axioms for a vector space are mere formalities, amounting to appending “$+K$” to each of the axioms for V.

Definition 4.56. Let $(V, +, \cdot)$ be a vector space, K a subspace. The set of cosets of K in V is denoted V/K. For $\boldsymbol{x}$, $\boldsymbol{y}$ in V and for real c, we define

$$(\boldsymbol{x} + K) \boxplus (\boldsymbol{y} + K) = (\boldsymbol{x} + \boldsymbol{y}) + K, \qquad c \boxdot (\boldsymbol{x} + K) = (c\boldsymbol{x}) + K.$$

The vector space $(V/K, \boxplus, \boxdot)$ is called the *quotient* of V by K.

Theorem 4.57. *If $(V, +, \cdot)$ is a vector space of dimension n, and if K is a subspace of dimension k, the quotient $(V/K, \boxplus, \boxdot)$ is a vector space of dimension $(n - k)$.*

Proof. Pick a basis $(\boldsymbol{v}_j)_{j=1}^k$ for K, and extend to a basis S of V by appending vectors $(\boldsymbol{v}_j)_{j=k+1}^n$. Write $V' = \mathrm{Span}(\boldsymbol{v}_j)_{j=k+1}^n$, and note that $V = K \oplus V'$. It suffices to show that the cosets $(\boldsymbol{v}_j + K)_{j=k+1}^n$ are a basis of V/K.

(Spanning). If $\boldsymbol{x} + K \in V/K$, then $\boldsymbol{x}$ can be written as a linear combination from S, say

$$\boldsymbol{x} = \sum_{j=1}^n x^j \boldsymbol{v}_j = \sum_{j=1}^k x^j \boldsymbol{v}_j + \sum_{j=k+1}^n x^j \boldsymbol{v}_j.$$

Calling the sums $\boldsymbol{x}_0$ and $\boldsymbol{x}'$, respectively, we have $\boldsymbol{x}_0 \in K$, so that $\boldsymbol{x} \equiv \boldsymbol{x}' \pmod{K}$, and $\boldsymbol{x}' \in V'$. As cosets (i.e., as elements of V/K), we have

$$\boldsymbol{x} + K = \boldsymbol{x}' + K = \sum_{j=k+1}^{n} x^j(\boldsymbol{v}_j + K).$$

(Linear independence). Suppose $x^{k+1}, \dots, x^n$ are scalars such that

$$\mathbf{0}^{V/K} = \sum_{j=k+1}^{n} x^j(\boldsymbol{v}_j + K) = \left(\sum_{j=k+1}^{n} x^j \boldsymbol{v}_j\right) + K.$$

The expression $\boldsymbol{x}_0$ in parentheses lies in V' by definition, and lies in K because $\boldsymbol{x}_0 + K = \mathbf{0}^{V/K}$. Since $K \cap V' = \{\mathbf{0}^V\}$, we have $\boldsymbol{x}_0 = \mathbf{0}^V$. Since $(\boldsymbol{v}_j)_{j=k+1}^n$ is linearly independent, $x^j = 0$ for each j. □

Theorem 4.58. *Let V and W be vector spaces. If $T : V \to W$ is a linear transformation with kernel K, there is a vector space isomorphism $\overline{T} : V/K \to T(V)$ defined by $\overline{T}(\boldsymbol{x} + K) = T(\boldsymbol{x})$.*

Proof. ($\overline{T}$ is well-defined). If $\boldsymbol{x}_1 + K = \boldsymbol{x}_2 + K$ for some $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ in V, then $\boldsymbol{x}_2 - \boldsymbol{x}_1 \in K = \ker(T)$, so $\mathbf{0}^W = T(\boldsymbol{x}_2 - \boldsymbol{x}_1) = T(\boldsymbol{x}_2) - T(\boldsymbol{x}_1)$, or $T(\boldsymbol{x}_1) = T(\boldsymbol{x}_2)$. That is, $\overline{T}(\boldsymbol{x}_1 + K) = \overline{T}(\boldsymbol{x}_2 + K)$.

(Injectivity). $\overline{T}(\boldsymbol{x}_0 + K) = T(\boldsymbol{x}_0) = \mathbf{0}^W$ if and only if $\boldsymbol{x}_0 \in K = \ker(T)$, if and only if $\boldsymbol{x}_0 + K = \mathbf{0}^{V/K}$.

(Surjectivity). Suppose $\boldsymbol{y} \in T(V)$. By hypothesis, there exists an $\boldsymbol{x}$ in V such that $T(\boldsymbol{x}) = \boldsymbol{y}$. By definition, $\overline{T}(\boldsymbol{x} + K) = T(\boldsymbol{x}) = \boldsymbol{y}$. □

Corollary 4.59 (The Rank-Nullity Theorem)**.** *Suppose V is a finite-dimensional vector space, W an arbitrary vector space. If $T : V \to W$ is a linear transformation, then*

$$\dim \ker(T) + \dim T(V) = \dim V.$$

Proof. Theorem 4.58 says $T(V)$ is isomorphic to $V/\ker(T)$. Theorem 4.57 immediately implies the claim. □

Corollary 4.60. *Let V and W be vector spaces of dimension n. A linear transformation $T : V \to W$ is injective if and only if it is surjective.*

Proof. By the Rank-Nullity Theorem, $\dim V = \dim\ker(T) + \dim T(V)$. By hypothesis, $\dim V = \dim W$. Thus, T is injective if and only if $\dim\ker(T) = 0$, if and only if $\dim T(V) = \dim V = \dim W$, if and only if $T(V) = W$, if and only if T is surjective. □

Remark 4.61. The conclusion of the Rank-Nullity Theorem is true if $\dim V = \infty$, in the sense that at least one of the kernel and the image is infinite-dimensional.

Remark 4.62. Let A be a real $m \times n$ matrix, $\mu_A : \mathbf{R}^n \to \mathbf{R}^m$ the multiplication map.

The dimension of $\mathbf{R}^n$ is the number of columns of A.

The kernel of μ_A is the solution space of the homogeneous system $A\boldsymbol{x} = \mathbf{0}^m$, i.e., the number of free variables after the coefficient matrix is put into row-echelon form.

A vector $\boldsymbol{b}$ in $\mathbf{R}^m$ is in the image of μ_A if and only if $A\boldsymbol{x} = \boldsymbol{b}$ is consistent. The solution set is a coset of the homogeneous solution space. This recapitulates our earlier result that an arbitrary non-homogeneous solution $\boldsymbol{x}$ is the sum of any particular solution $\boldsymbol{x}_0$ and a homogeneous solution $\boldsymbol{x}_h$.

By the Rank-Nullity Theorem, the dimension of the image is the number of columns minus the number of free variables, i.e., the number of basic variables.

Exercises

Exercise 4.1. In the vector space P_3 of cubic polynomials, consider the basis $S'' = (1, 1+t, 1+t+\frac{1}{2}t^2, 1+t+\frac{1}{2}t^2+\frac{1}{6}t^3) = (\boldsymbol{v}_j'')_{j=1}^4$. Use the techniques (and results, as appropriate) of Example 4.42 to calculate $[D]_{S''}^{S''}$.

Exercise 4.2. Let V be the vector space P_2 of quadratic polynomials, and let $W = \mathbf{R}^3$, both with the usual vector space operations. Define $T : V \to W$ by $T(p) = \big(p(-1), p(0), p(1)\big)$.

(a) Show directly that T is linear.

(b) Find the matrix of T with respect to the standard bases.

(c) Determine (with justification) whether or not T is an isomorphism.

(d) If possible, solve $T(p) = (1, 2, 3)$ for p.

Exercise 4.3. Each part refers to the matrix

$$A = \begin{bmatrix} -1 & 2 & -3 \\ -2 & 4 & -6 \\ 1 & -2 & 3 \end{bmatrix}$$

and the associated linear transformaton $\mu_A : \mathbf{R}^3 \to \mathbf{R}^3$.

(a) Use row reduction to find a basis for $\ker(\mu_A)$, and for $\operatorname{im}(\mu_A)$. Suggestion: Use Rank-Nullity to find $\dim \operatorname{im}(\mu_A)$.

(b) Find vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbf{R}^3$ such that $A = \boldsymbol{w}\boldsymbol{v}^{\mathsf{T}}$, and verify that $\ker(\mu_A) = \operatorname{Span}(\boldsymbol{v})^{\perp}$ and $\operatorname{im}(\mu_A) = \operatorname{Span}(\boldsymbol{w})$.

Exercise 4.4. Let $S = (\boldsymbol{e}_1^1, \boldsymbol{e}_1^2, \boldsymbol{e}_2^1, \boldsymbol{e}_2^2)$ be the standard basis of $\mathbf{R}^{2\times 2}$. Calculate the matrices with respect to S of (a) The transpose operator; (b) The operator $T(A) = A + A^{\mathsf{T}}$; (c) The operator $T(A) = A - A^{\mathsf{T}}$.

Exercise 4.5. Let $V \subseteq \mathcal{F}(\mathbf{R}, \mathbf{R})$ be the subspace spanned by the set $S = (\cos, \sin) = (\boldsymbol{v}_j)_{j=1}^2$.

(a) Show that differentiation defines an operator $D : V \to V$, and find the matrix $J = [D]_S^S$.

(b) Calculate J^2, the square of the matrix you found in (a), and interpret the result in terms of second derivatives of cos and sin.

Exercise 4.6. Let n be a positive integer.

(a) Prove that P_n is isomorphic (as a vector space) to $\mathbf{R}^{n+1}$. Suggestion: There is an "obvious" way to map a polynomial of degree at most n to an ordered tuple of coefficients.

(b) Let t_0 be a real number, and define the "evaluation functional" $\varepsilon_{t_0} : P_n \to \mathbf{R}$ by $\varepsilon_{t_0}(p) = p(t_0)$. Use the Rank-Nullity Theorem to prove that $\ker(T) = \{p \text{ in } P_n : p(t_0) = 0\}$ is n-dimensional.

(c) In the notation of part (b), show that $S = \big((t - t_0)t^j\big)_{j=0}^{n-1}$ is a basis of $\ker(T)$.

Exercise 4.7. Let a, b, and c be real numbers, not all 0,

$$A = \begin{bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{bmatrix},$$

and let $\mu_A : \mathbf{R}^3 \to \mathbf{R}^3$ be multiplication by A.

(a) Show that $\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = -\langle \boldsymbol{x}, A\boldsymbol{y}\rangle$ for all $\boldsymbol{x}$, $\boldsymbol{y}$ in $\mathbf{R}^3$.
Hint: Recall that $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}$.

(b) Find the rank of μ_A.
Suggestion: Split into cases (i) $a \neq 0$; (ii) $a = 0$ and $b \neq 0$; and (iii) $a = b = 0$, $c \neq 0$.

(c) Find the nullity of μ_A, give a basis for the kernel, and use part (a) to show that $\mathbf{R}^3$ is the *orthogonal* direct sum $\ker(\mu_A) \oplus \operatorname{im}(\mu_A)$.

Exercise 4.8. Prove from the definition that a composition of linear transformations is linear.

Exercise 4.9. Let $T : V \to V'$ and $T' : V' \to V''$ be linear transformations. Prove:

(a) $\ker(T) \subseteq \ker(T'T)$.

(b) $\operatorname{im}(T'T) \subseteq \operatorname{im}(T')$.

Exercise 4.10. Suppose A, B, and P are $n \times n$ matrices, and that $B = P^{-1}AP$. Prove by mathematical induction that if N is a positive integer, then $B^N = P^{-1}A^NP$. (Note carefully that this is not what you might expect by "distributing" the exponent!)

Exercise 4.11. Let a_1 and a_2 be real numbers, and define

$$A = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_1 \end{bmatrix}, \qquad \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

(a) Show $A\boldsymbol{v}_1 = (a_1 + a_2)\boldsymbol{v}_1$. Establish a similar formula for $A\boldsymbol{v}_2$.

(b) Let $S = (\boldsymbol{e}_1, \boldsymbol{e}_2)$ be the standard basis of $\mathbf{R}^2$, and $S' = (\boldsymbol{v}_1, \boldsymbol{v}_2)$. Calculate the transition matrices $[I]_S^{S'}$ and $[I]_{S'}^{S}$.

(c) Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear transformation whose standard matrix $[T]_S^S$ is A. Find the matrix $[T]_{S'}^{S'}$.

(d) If N is a positive integer, find a formula for the entries of A^N.
Suggestion: Use Exercise 4.10.

Exercise 4.12. Let σ be a permutation (i.e., bijection) of the set $\{1, 2, \dots, n\}$, and let P_σ be the associated permutation matrix, see Exercise 1.13. If $A = \operatorname{diag}[\lambda^1, \lambda^2, \dots, \lambda^n]$ for some numbers λ^i, calculate $\mathsf{P}_\sigma A \mathsf{P}_\sigma^{-1}$.

Exercise 4.13. In each part, let V be the space of smooth (infinitely-differentiable) real-valued functions on $\mathbf{R}$, let V_0 be the subspace of functions f satisfying $f(0) = 0$, and let D and S denote the differentiation and integration operators

$$(Df) = f'(t), \qquad (Sf)(t) = \int_0^t f(s)\,ds.$$

Use theorems from calculus to justify your answer in each part. (Exercise 4.9 may also be helpful.)

(a) Calculate the compositions $D \circ S$ and $S \circ D$. Is either composition the identity map I_V?

(b) Show S is injective, and find the image.

(c) Show D is surjective, and find the kernel.

(d) Does S map V_0 to itself? If so, is $S : V_0 \to V_0$ an isomorphism?

(e) Does D map V_0 to itself? If so, is $D : V_0 \to V_0$ an isomorphism?

Exercise 4.14. Let ℓ be a positive real number. A function $f : \mathbf{R} \to \mathbf{R}$ is *ℓ-periodic* if $f(t + \ell) = f(t)$ for all real t. With the notation of the preceding exercise:

(a) Prove that the space of ℓ-periodic functions is a vector subspace of $\mathcal{F}(\mathbf{R}, \mathbf{R})$. (Consequently, the set W of *smooth* ℓ-periodic functions is a subspace of V.)

(b) Show that D maps W to W. (That is, the derivative of a smooth, ℓ-periodic function is ℓ-periodic.)

(c) Show that if $f \in W$, then $Sf \in W$ if and only if $\int_0^\ell f(t)\,dt = 0$.

Exercise 4.15. Let V be a finite-dimensional vector space with basis $S = (\boldsymbol{v}_j)_{j=1}^n$, let $T : V \to W$ be a linear transformation, and put $\boldsymbol{w}_j = T(\boldsymbol{v}_j)$ for each j. Prove from the definitions (not using results proven in this chapter) that:

(a) The set $T(S) = (\boldsymbol{w}_j)_{j=1}^n$ spans the image $T(V)$.

(b) If $(\boldsymbol{v}_j)_{j=1}^k$ is a basis for $\ker(T)$, then the set $(\boldsymbol{w}_j)_{j=k+1}^n$ is a basis of the image $T(V)$.

(c) Use part (b) to deduce the Rank-Nullity Theorem (Corollary 4.59).

Chapter 5

Diagonalization

5.1 Linear Operators

If $(V, +, \cdot)$ is a vector space, a linear transformation $T : V \to V$ is called a *linear operator* (on V). If V is finite-dimensional, and if S and S' are bases, an operator T may be expressed as a matrix with respect to either basis, and the two matrices are similar:

$$[T]_{S'}^{S'} = [I_V]_S^{S'}[T]_S^S[I_V]_{S'}^S, \quad \text{or} \quad A' = P^{-1}AP.$$

Given a linear operator T, we would like to find a basis S' with respect to which the matrix A' of T is "as simple as possible".* For example, we might want A' to be a diagonal matrix or an upper triangular matrix. Can we arrange this, and if so, how can we calculate a basis S' for a specific operator?

Remark 5.1. If A' were a *scalar matrix*, i.e., if $A' = cI_n$ for some real number c, the condition $A' = P^{-1}AP$ would imply

$$A = PA'P^{-1} = P(cI_n)P^{-1} = cI_n,$$

see Theorem 1.19 (iii). Consequently, every matrix similar to a scalar matrix *is* a scalar matrix.

Definition 5.2. Let $(V, +, \cdot)$ be an n-dimensional real vector space. A linear operator $T : V \to V$ is *diagonalizable* if there exists a basis S' of V such that $A' = [T]_{S'}^{S'}$ is diagonal.

*The technical term is a *canonical form* for T, or for A. A matrix does not have a unique "canonical form"; the notion of "simplest" depends on one's mathematical intent.

The following characterization is immediate from the definition, but stated formally for emphasis.

Proposition 5.3. *Let $(V, +, \cdot)$ be an n-dimensional real vector space, $S' = (\boldsymbol{v}_j)_{j=1}^n$ a basis of V, and $T : V \to V$ a linear operator. The matrix $A' = [T]_{S'}^{S'}$ is diagonal if and only if there exist scalars $(\lambda^j)_{j=1}^n$ such that $T(\boldsymbol{v}_j) = \lambda^j \boldsymbol{v}_j$ for each j.*

Remark 5.4. In the notation of the proposition,

$$[T]_{S'}^{S'} = \operatorname{diag}[\lambda^1, \lambda^2, \dots, \lambda^n] = \operatorname{diag}[\lambda^j] = \begin{bmatrix} \lambda^1 & 0 & \dots & 0 \\ 0 & \lambda^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda^n \end{bmatrix}.$$

5.2 Eigenvalues and Eigenvectors

Definition 5.5. Let $T : V \to V$ be a linear operator. A real number λ is an *eigenvalue* of T if the operator $(T - \lambda I_V)$ is not invertible.

A non-zero vector $\boldsymbol{v}$ in V is a λ-*eigenvector* of T if $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$.

If λ is an eigenvalue of T, the subspace

$$E_\lambda = \ker(T - \lambda I_V) = \{v \text{ in } V : T(\boldsymbol{v}) = \lambda \boldsymbol{v}\}$$

is the λ-*eigenspace* of T.

Remark 5.6. If λ is an eigenvalue of T, the λ-eigenspace consists of all λ-eigenvectors together with the zero vector $\mathbf{0}^V$, which by definition is not an eigenvector but does satisfy $T(\boldsymbol{v}) = \lambda \boldsymbol{v}$.

Eigenspaces of an operator are "invariants", depending only on the operator, not on any choice of basis of V.

Every linear operator on a 1-dimensional space is scalar, hence diagonal with respect to an arbitrary basis.

Remark 5.7. Proposition 5.3 says T is diagonalizable if and only if V has a basis consisting entirely of eigenvectors of T. For brevity, such a basis is called a T-*eigenbasis* of V.

Example 5.8. If $A = \operatorname{diag}[\lambda^1, \lambda^2, \dots, \lambda^n]$, then every standard basis vector is an eigenvector: $A\boldsymbol{e}_j = \lambda^j \boldsymbol{e}_j$. If the λ^j are distinct, the eigenspaces of A are precisely the coordinate axes. If there are repetitions among the diagonal entries, the standard basis vectors corresponding to a single eigenvalue λ span the λ-eigenspace.

Example 5.9. A linear operator T has eigenvalue 0 if and only if T is not invertible, and $E_0 = \ker(T)$.

Example 5.10. Let a and b be real numbers, and let $T : \mathbf{R}^2 \to \mathbf{R}^2$ have standard matrix

$$[T]_S^S = A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

The vector $\boldsymbol{v}_1 = (1, 1)$ is easily checked to be an $(a + b)$-eigenvector, while $\boldsymbol{v}_2 = (1, -1)$ is an $(a - b)$-eigenvector. It follows that T is diagonalizable, and if $S' = (\boldsymbol{v}_j)_{j=1}^2$, then $[T]_{S'}^{S'} = A' = \operatorname{diag}[a + b, a - b]$. (Checking this by direct computation is worthwhile, see Exercise 4.11.)

If $b \neq 0$, the eigenvalues $a \pm b$ of T are distinct, and there are two 1-dimensional eigenspaces:

$$E_{a+b} = \operatorname{Span}(\boldsymbol{v}_1), \qquad E_{a-b} = \operatorname{Span}(\boldsymbol{v}_2).$$

If $b = 0$, there is one eigenvalue, a, and $E_a = \mathbf{R}^2$.

Example 5.11. Let P denote the space of polynomials in one variable, and D the derivative operator.

A non-zero constant polynomial is a 0-eigenvector for D. There are no other eigenvectors in P: If p is a polynomial of degree $n \geq 1$, then $D(p) = p'$ is a non-zero polynomial of degree $n - 1 \geq 0$, while every scalar multiple of p is either zero or of degree n.

The operator $T(p)(t) = tp'(t)$ on P has each positive integer k as eigenvalue, since $T(t^k) = t(kt^{k-1}) = kt^k$. Since these eigenvectors span P, it turns out there are no other eigenvalues or eigenspaces.

Remark 5.12. If $T : V \to V$ is an operator and c is real, then λ is an eigenvalue of T if and only if $(\lambda - c)$ is an eigenvalue of $T - cI_V$.

Theorem 5.13. *Let $(V, +, \cdot)$ be an n-dimensional real vector space, $T : V \to V$ a linear operator.*

(i) *If λ is an eigenvalue of T, then* $\dim E_\lambda \geq 1$.

(ii) *If $\lambda^1 > \lambda^2 > \cdots > \lambda^r$ are eigenvalues of T, and if $S = (\boldsymbol{v}_j)_{j=1}^r$ is a set of corresponding eigenvectors, i.e., if $T(\boldsymbol{v}_j) = \lambda^j \boldsymbol{v}_j$ for each j, then S is linearly independent.*

(iii) *T is diagonalizable if and only if its eigenspaces span V.*

Proof. (i). If λ is an eigenvalue of T, then $T - \lambda I_V$ is not invertible. By Corollary 4.60, $T - \lambda I_V$ is not injective, so $\dim\ker(T - \lambda I_V) \geq 1$. (Concretely, there is a non-zero vector $\boldsymbol{v}$ such that $(T - \lambda I_V)(\boldsymbol{v}) = \mathbf{0}^V$, i.e., $T(\boldsymbol{v}) = \lambda\boldsymbol{v}$.)

(ii). The proof is by induction on r. The case $r = 1$ is obvious: an eigenvector is non-zero by definition, so a set of one eigenvector is linearly independent. Assume inductively for some $k \geq 1$ that the set $(\boldsymbol{v}_j)_{j=1}^k$ of eigenvectors is linearly independent, and suppose

$$(*) \qquad \mathbf{0}^V = x^1\boldsymbol{v}_1 + \cdots + x^k\boldsymbol{v}_k + x^{k+1}\boldsymbol{v}_{k+1} = \sum_{j=1}^{k+1} x^j\boldsymbol{v}_j$$

for some scalars x^j. Applying the operator $T - \lambda^{k+1}I_V$ to the preceding equation gives

$$\mathbf{0}^V = x^1(\lambda^1 - \lambda^{k+1})\boldsymbol{v}_1 + \cdots + x^k(\lambda^k - \lambda^{k+1})\boldsymbol{v}_k = \sum_{j=1}^{k} x^j(\lambda^j - \lambda^{k+1})\boldsymbol{v}_j.$$

The right-hand sum is a linear combination from a linearly independent set, so the coefficients $x^j(\lambda^j - \lambda^{k+1})$ are all 0. By hypothesis, the eigenvalues are decreasing with index, so $0 < (\lambda^j - \lambda^{k+1})$; thus $x^j = 0$ for $1 \leq j \leq k$. Substituting into (*), we find $x^{k+1} = 0$ as well; that is, $(\boldsymbol{v}_j)_{j=1}^{k+1}$ is linearly independent.

(iii). As noted above, T is diagonalizable if and only if there exists a T-eigenbasis of V.

Suppose V has a T-eigenbasis S'. Since every eigenvector of T lies in some eigenspace, $S' \subseteq \bigcup_\lambda E_\lambda$, and therefore

$$V = \operatorname{Span}(S') \subseteq \operatorname{Span}\Bigl(\bigcup_\lambda E_\lambda\Bigr) = \bigoplus_\lambda E_\lambda.$$

(The final step is Theorem 2.44; the sum is direct by (ii) just proven.)

Conversely, assume the eigenspaces of T span V. For each eigenvalue λ of T, pick a basis $S_\lambda = (\boldsymbol{v}_j^\lambda)_{j=1}^{n_\lambda}$ of E_λ. It suffices to prove the union of these sets, $S = \bigcup_\lambda S_\lambda$, is a basis of V.

Since $\operatorname{Span}(S_\lambda) = E_\lambda$, Theorem 2.44 gives

$$\operatorname{Span}(S) = \bigoplus_\lambda E_\lambda = V.$$

To prove S is linearly independent, suppose some linear combination from S is equal to $\mathbf{0}^V$. Indexing coefficients (and grouping terms) according to the corresponding eigenvalue,

$$\mathbf{0}^V = \sum_{\lambda}\left(\sum_{j=1}^{n_\lambda} x_\lambda^j \boldsymbol{v}_j^\lambda\right).$$

This equation has the form $\mathbf{0}^V = \sum_\lambda \boldsymbol{v}^\lambda$, with $\boldsymbol{v}^\lambda$ in E_λ. By (ii), $\boldsymbol{v}^\lambda = \mathbf{0}^V$; if any $\boldsymbol{v}^\lambda$ were non-zero, the non-trivial linear combination $\sum_\lambda \boldsymbol{v}^\lambda$ would be non-zero. Fix an eigenvalue λ arbitrarily. Since

$$\sum_{j=1}^{n_\lambda} x_\lambda^j \boldsymbol{v}_j^\lambda = \boldsymbol{v}^\lambda = \mathbf{0}^V$$

and $S_\lambda = (\boldsymbol{v}_j^\lambda)_{j=1}^{n_\lambda}$ is linearly independent, all the scalars x_λ^j are zero; only the trivial linear combination from S gives the zero vector. □

Corollary 5.14. *If* $\dim V = n$ *and* T *has* n *distinct real eigenvalues, then* T *is diagonalizable.*

Proof. For each eigenvalue λ there exists a λ-eigenvector by (i); these n vectors are linearly independent by (ii), so are a basis of V. □

Example 5.15. If $\boldsymbol{v}$ and $\boldsymbol{w}$ are arbitrary non-zero elements of $\mathbf{R}^n$, there is a rank-one operator $T(\boldsymbol{x}) = \langle \boldsymbol{v}, \boldsymbol{x}\rangle \cdot \boldsymbol{w}$, with standard matrix $\boldsymbol{w}\boldsymbol{v}^{\mathsf{T}}$, see Example 4.31.

If $\langle \boldsymbol{v}, \boldsymbol{x}\rangle = 0$, then $T(\boldsymbol{x}) = \mathbf{0}^n$; that is, the 0-eigenspace contains $\mathrm{Span}(\boldsymbol{v}^\perp)$, which has dimension $(n-1)$. (The 0-eigenspace is not $\mathbf{R}^n$, because $\boldsymbol{w}\boldsymbol{v}^{\mathsf{T}} \neq \mathbf{0}^{n\times n}$, so in fact $E_0 = \mathrm{Span}(\boldsymbol{v}^\perp)$.)

If $\boldsymbol{x}$ is an eigenvector of T and $\langle \boldsymbol{v}, \boldsymbol{x}\rangle \neq 0$, then $\langle \boldsymbol{v}, \boldsymbol{x}\rangle \cdot \boldsymbol{w} = \lambda \boldsymbol{x}$. It follows that $\boldsymbol{x} = c\boldsymbol{w}$ with $c \neq 0$, in which case $\lambda = \langle \boldsymbol{v}, \boldsymbol{w}\rangle$.

If $\langle \boldsymbol{v}, \boldsymbol{w}\rangle \neq 0$, then $E_\lambda = \mathrm{Span}(\boldsymbol{w})$. Consequently, $\mathbf{R}^n = E_0 \oplus E_\lambda$, and T is diagonalizable. If instead $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0$, then all n eigenvalues of T are 0 but $T \neq \mathbf{0}$, so T is not diagonalizable.

Involutions

Definition 5.16. Let V be an arbitrary vector space (possibly infinite-dimensional). An operator T satisfying $T^2 = I_V$ is an *involution* of V.

Proposition 5.17. *If* T *is an involution, the eigenspaces of* T *span* V.

Proof. If $\boldsymbol{v}$ is an arbitrary element of V, then the vectors

$$\boldsymbol{v}^1 = \tfrac{1}{2}\big(\boldsymbol{v} + T(\boldsymbol{v})\big), \qquad \boldsymbol{v}^{-1} = \tfrac{1}{2}\big(\boldsymbol{v} - T(\boldsymbol{v})\big),$$

are eigenvectors of T with eigenvalue 1 and -1, respectively (compare the proof of Proposition 2.49). Moreover, $\boldsymbol{v} = \boldsymbol{v}^1 + \boldsymbol{v}^{-1}$; that is, the eigenspaces of T span V. □

Remark 5.18. The eigenvalues of an involution are, *a priori*, ± 1, since if $\boldsymbol{v}$ is an eigenvector, then

$$\boldsymbol{v} = T^2(\boldsymbol{v}) = T\big(T(\boldsymbol{v})\big) = T(\lambda\boldsymbol{v}) = \lambda T(\boldsymbol{v}) = \lambda^2\boldsymbol{v}.$$

Suppose "optimistically" that every vector decomposes as a sum of eigenvectors: $\boldsymbol{v} = \boldsymbol{v}^1 + \boldsymbol{v}^{-1}$. Applying T gives $T(\boldsymbol{v}) = \boldsymbol{v}^1 - \boldsymbol{v}^{-1}$. "Solving" for $\boldsymbol{v}^1$ and $\boldsymbol{v}^{-1}$ gives the formulas in the preceding proof.

Example 5.19. Let $V = \mathbf{R}^{n\times n}$. The transpose operator $T(A) = A^{\mathsf{T}}$ is an involution (the transpose of a transpose is the original matrix). Its eigenspaces are $E_1 = \mathrm{Sym}^n$ (the symmetric matrices) and $E_{-1} = \mathrm{Skew}^n$ (the skew-symmetric matrices), compare Proposition 2.49.

5.3 The Characteristic Polynomial

Throughout this section, $(V, +, \cdot)$ is a finite-dimensional real vector space of dimension n, and $T : V \to V$ is a linear operator.

Proposition 5.20. *If S and S' are bases of V, and if $A = [T]_S^S$ and $A' = [T]_{S'}^{S'}$ are the matrices of T in these bases, then* $\det A' = \det A$.

Proof. By Corollary 4.37 (ii), the matrices A and A' are similar. By Corollary 3.57, similar matrices have the same determinant. □

Definition 5.21. The *determinant* of T is the determinant of the matrix of T with respect to an arbitrary basis of V.

The *characteristic polynomial* of T is

$$\chi_T(\lambda) = \det(T - \lambda I_V) = \det T + \cdots + (-\lambda)^n,$$

of degree n. The *characteristic equation* of T is $\chi_T(\lambda) = 0$.

Remark 5.22. Each summand in a determinant is a signed product of n matrix entries. Since each entry in the matrix of $T - \lambda I_V$ is either a number or a linear polynomial in λ, the determinant has degree n. The only term of degree n comes from the product of the diagonal entries $(A_j^j - \lambda)$, so the coefficient of λ^n is $(-1)^n$. The constant term is found by setting $\lambda = 0$.

Proposition 5.23. *A real number λ is an eigenvalue of T if and only if $\chi_T(\lambda) = 0$.*

Proof. A real number λ is an eigenvalue of T if and only if $T - \lambda I_V$ is not invertible, if and only if $\chi_T(\lambda) = \det(T - \lambda I_V) = 0$. □

Remark 5.24. By the Fundamental Theorem of Algebra, the characteristic polynomial of T has precisely n complex roots, counting multiplicity. For terminological convenience, we extend the definition of "eigenvalue" to encompass non-real roots.

Definition 5.25. Let $(V, +, \cdot)$ be a finite-dimensional real vector space, and $T : V \to V$ a linear operator. A root of the characteristic polynomial of T is an *eigenvalue* of T.

Example 5.26. Let a and b be real, and consider the matrix

$$A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}; \quad \chi_A(\lambda) = \det(A - \lambda I_2) = \begin{vmatrix} a - \lambda & b \\ b & a - \lambda \end{vmatrix} = (a - \lambda)^2 - b^2,$$

a difference of squares: $\chi_A(\lambda) = (a + b - \lambda)(a - b - \lambda)$. We recover the eigenvalues $\lambda = a \pm b$. (Computing corresponding eigenvectors by substituting each eigenvalue in turn and solving the resulting homogeneous system is a worthwhile exercise.)

Example 5.27. Let a and b be real, and consider the matrix

$$B = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}; \quad \chi_B(\lambda) = \det(B - \lambda I_2) = \begin{vmatrix} a - \lambda & -b \\ b & a - \lambda \end{vmatrix} = (a - \lambda)^2 + b^2.$$

The characteristic equation of B is $\lambda^2 - 2a\lambda + (a^2 + b^2) = 0$. The quadratic formula gives

$$\lambda = \frac{2a \pm \sqrt{(2a)^2 - 4(a^2 + b^2)}}{2} = a \pm \sqrt{-b^2} = a \pm bi.$$

That is, if $b \neq 0$, the roots of the characteristic equation are non-real. Because B has no real eigenvalues, B has no real eigenvectors. It follows that B is not diagonalizable *over the real numbers*.

Remark 5.28. The matrix B of the preceding example defines a linear operator on the complex vector space $\mathbf{C}^2$, the set of ordered pairs of complex numbers equipped with *complex* scalar multiplication. In this space, B has two linearly independent eigenvectors: $\boldsymbol{v}_1 = (1, -i)$ and $\boldsymbol{v}_2 = (1, i)$. For example,

$$B\boldsymbol{v}_1 = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} a+bi \\ b-ai \end{bmatrix} = (a+bi) \begin{bmatrix} 1 \\ -i \end{bmatrix} = (a+bi)\boldsymbol{v}_1.$$

Thus, B is diagonalizable *over the complex numbers*.

Example 5.29. Let a and $b \neq 0$ be real, and consider the matrix

$$C = \begin{bmatrix} a & b \\ 0 & a \end{bmatrix}; \quad \chi_C(\lambda) = \det(C - \lambda I_2) = \begin{vmatrix} a-\lambda & b \\ 0 & a-\lambda \end{vmatrix} = (a-\lambda)^2.$$

The only real eigenvalue of C is a, of "algebraic multiplicity 2". (That is, $\lambda = a$ is a double root of the characteristic equation.)

To find the corresponding eigenspace, we solve the homogeneous system with coefficient matrix $C - aI_2 = b\boldsymbol{e}_1^2$. Multiplying the first row by $1/b$ puts the coefficient matrix into reduced row-echelon form. The variable x^2 is basic, and x^1 is free. The only eigenvectors are therefore scalar multiples of $\boldsymbol{v}_1 = (1, 0)$, i.e.,

$$E_a = \{(x^1, 0) \text{ in } \mathbf{R}^2\}.$$

The eigenspace does not span $\mathbf{R}^2$, so C is not diagonalizable. Because $\dim E_a = 1$, we say the eigenvalue a has "geometric multiplicity 1".

Remark 5.30. The preceding three examples typify the general eigenspace behavior of a real $n \times n$ matrix A. In the "best" case, A has all real eigenvalues and a set of eigenvectors that span $\mathbf{R}^n$, so A is diagonalizable.

Two things can prevent A from being diagonalizable over the reals. First, some of the eigenvalues may be non-real. (Because the characteristic polynomial of a real matrix has real coefficients, any non-real eigenvalues occur in complex conjugate pairs. A matrix with non-real eigenvalues *may* be diagonalizable over the complex numbers, i.e., there may exist an invertible *complex* $n \times n$ matrix P such that $P^{-1}AP$ is diagonal.) Second, A may have only real eigenvalues, but its eigenspaces may fail to span $\mathbf{R}^n$.

Example 5.31. Find the eigenvalues and eigenspaces of the matrix

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix},$$

and determine whether A is diagonalizable. If so, find a matrix P so that $A' = P^{-1}AP$ is diagonal with the eigenvalues in non-increasing order.

Subtract λI_3, then expand the determinant using the formula from Example 3.47:

$$\begin{aligned}
\det(A - \lambda I_3) &= \begin{vmatrix} 1-\lambda & 1 & -1 \\ 1 & 1-\lambda & 1 \\ 1 & -1 & 3-\lambda \end{vmatrix} \\
&= (1-\lambda)\big[(1-\lambda)(3-\lambda) + 1\big] \\
&\qquad - \big[(3-\lambda) - 1\big] \\
&\qquad + (-1)\big[(1)(-1) - (1-\lambda)\big] \\
&= (1-\lambda)\big[\lambda^2 - 4\lambda + 4\big] + \big[\lambda - 2\big] - \big[\lambda - 2\big] \\
&= -(\lambda - 1)(\lambda - 2)^2,
\end{aligned}$$

so the eigenvalues (in non-increasing order) are 2, 2, and 1.

To find a basis for E_2, row-reduce $A - 2I_3$:

$$A - 2I_3 = \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \xrightarrow{\substack{R_3 - R_2, \\ R_1 + R_2}} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The system may be written $x^1 = x^2 - x^3$, so the general solution is

$$\begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} x^2 - x^3 \\ x^2 \\ x^3 \end{bmatrix} = x^2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x^3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}; \quad \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},\ \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Theorem 5.13 implies A is diagonalizable: We have shown $\dim(E_2) = 2$, and $\dim(E_1) \geq 1$, so the eigenspaces of A span $\mathbf{R}^3$. To *find* a 1-eigenvector, row-reduce $A - I_3$:

$$A - I_3 = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix} \xrightarrow{\substack{R_3 - R_2, \\ R_3 + R_1}} \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The system is written $x^1 = -x^3$ and $x^2 = x^3$, so

$$\begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} -x^3 \\ x^3 \\ x^3 \end{bmatrix} = x^3 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}; \quad \boldsymbol{v}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

To catch arithmetic mistakes, it's a good idea to verify by matrix multiplication that $A\boldsymbol{v}_1 = 2\boldsymbol{v}_1$, $A\boldsymbol{v}_2 = 2\boldsymbol{v}_2$, and $A\boldsymbol{v}_3 = \boldsymbol{v}_3$.

The eigenbasis $S' = (\boldsymbol{v}_j)_{j=1}^3$ diagonalizes. The transition matrix $P = [I_3]_{S'}^S$ has jth column $[\boldsymbol{v}_j]^S$. To compute the inverse, $P^{-1} = [I_3]_S^{S'}$, form the augmented matrix $\begin{bmatrix} P \mid I_3 \end{bmatrix}$ and row-reduce:

$$\left[\begin{array}{rrr|rrr} 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\substack{R_1+R_3,\\ R_2-R_1,\\ R_3-R_2}} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 1 & -1 \\ 0 & 1 & 0 & 1 & -1 & 2 \end{array}\right];$$

swapping the second and third rows gives

$$P^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 2 \\ -1 & 1 & -1 \end{bmatrix}, \qquad P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Again to guard against arithmetic errors, it's a good idea to check that $P^{-1}P = I_3$. On theoretical grounds, $\operatorname{diag}[2, 2, 1] = A' = P^{-1}AP$, i.e.,

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 2 \\ -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

For good measure, check this by multiplying matrices.

Example 5.32. Let $k \geq 2$ be an integer and λ real. The $k \times k$ matrix

$$B_k(\lambda) = \lambda I_k + \sum_{j=1}^{k-1} e_j \boldsymbol{e}^{j+1} = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix}$$

is called the *Jordan block* of size k with eigenvalue λ. The matrix $N_k = B_k(0)$ is *nilpotent*: $N_k^j \neq \mathbf{0}^{k\times k}$ for $1 \leq j < k$, but $N_k^k = \mathbf{0}^{k\times k}$.

The eigenvalue of $B_k(\lambda)$ is λ, with algebraic multiplicity k. The λ-eigenspace is easily checked to be $\mathrm{Span}(\boldsymbol{e}_1)$, so λ has geometric multiplicity 1. In particular, $B_k(\lambda)$ is not diagonalizable.

In more advanced courses, it is proven that if V is finite-dimensional and T is an operator with real eigenvalues, then V splits as a direct sum of subspaces in such a way that T acts as a Jordan block on each subspace.

5.4 Symmetric Operators

Let N be a positive integer. Throughout this section, V is a subspace of $(\mathbf{R}^N, +, \cdot)$, and is equipped with the inner product coming from the Euclidean dot product, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\mathsf{T} \boldsymbol{y}$.

Definition 5.33. An operator $T : V \to V$ is *symmetric* if

$$\langle T(\boldsymbol{x}), \boldsymbol{y} \rangle = \langle \boldsymbol{x}, T(\boldsymbol{y}) \rangle \quad \text{for all } x \text{ and } y \text{ in } V.$$

Remark 5.34. If $(\boldsymbol{v}_j)_{j=1}^n$ is a basis of V, T is symmetric if and only if

$$\langle T(\boldsymbol{v}_i), \boldsymbol{v}_j \rangle = \langle \boldsymbol{v}_i, T(\boldsymbol{v}_j) \rangle \quad \text{for all } i,\ j = 1,\ \dots,\ n.$$

Symmetry implies this condition *a fortiori*. Conversely, if this condition holds, then writing $\boldsymbol{x} = \sum_i x^i \boldsymbol{v}_i$ and $\boldsymbol{y} = \sum_j y^j \boldsymbol{v}_j$ and using bilinearity of the dot product shows T is symmetric.

Example 5.35. Let $V = \mathbf{R}^n$, and assume $T(\boldsymbol{x}) = A\boldsymbol{x}$ for all $\boldsymbol{x}$ in V; that is, A is the standard matrix of T. The operator T is symmetric if and only if

$$\boldsymbol{x}^\mathsf{T} A^\mathsf{T} \boldsymbol{y} = (A\boldsymbol{x})^\mathsf{T} \boldsymbol{y} = \langle A\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, A\boldsymbol{y} \rangle = \boldsymbol{x}^\mathsf{T} A \boldsymbol{y}$$

for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^n$, if and only if $A^\mathsf{T} = A$.

Proposition 5.36. *If W is a subspace of $(\mathbf{R}^n, +, \cdot)$, the orthogonal projection map $\mathrm{proj}_W : \mathbf{R}^n \to \mathbf{R}^n$ is symmetric.*

Proof. Every vector $\boldsymbol{x}$ in $\mathbf{R}^n$ decomposes *uniquely* as $\boldsymbol{x} = \boldsymbol{x}_W + \boldsymbol{x}_{W^\perp}$, and the orthogonal projection operator is defined by $\mathrm{proj}_W(\boldsymbol{x}) = \boldsymbol{x}_W$.

If $\boldsymbol{x} = \boldsymbol{x}_W + \boldsymbol{x}_{W^\perp}$ and $\boldsymbol{y} = \boldsymbol{y}_W + \boldsymbol{y}_{W^\perp}$ are arbitrary vectors, then

$$\begin{aligned} \langle \mathrm{proj}_W(\boldsymbol{x}), \boldsymbol{y} \rangle &= \langle \boldsymbol{x}_W, \boldsymbol{y}_W + \boldsymbol{y}_{W^\perp} \rangle \\ &= \langle \boldsymbol{x}_W, \boldsymbol{y}_W \rangle \\ &= \langle \boldsymbol{x}_W + \boldsymbol{x}_{W^\perp}, \boldsymbol{y}_W \rangle = \langle \boldsymbol{x}, \mathrm{proj}_W(\boldsymbol{y}) \rangle . \end{aligned}$$ □

Proposition 5.37. *If T_1 and T_2 are symmetric, and if c is real, then $cT_1 + T_2$ is symmetric.*

Proof. Exercise 5.3. □

Remark 5.38. In particular, the set of symmetric operators is a vector subspace of the set $\mathcal{L}(V, V)$ of all operators. Further, an arbitrary linear combination of orthogonal projections is symmetric; the *Spectral Theorem*, below, asserts the converse.

Proposition 5.39. *Let $S' = (\boldsymbol{u}_j)_{j=1}^n$ be an* orthonormal *basis of V. If T is an operator on V, then T is symmetric if and only if the matrix $[T]_{S'}^{S'}$ is a symmetric matrix.*

Proof. The matrix $A = [A_j^i]$ of T with respect to S' is defined by

$$T(\boldsymbol{u}_j) = \sum_{k=1}^n A_j^k \boldsymbol{u}_k.$$

Since $\langle \boldsymbol{u}_i, \boldsymbol{u}_k \rangle = \delta_{ik}$, we have

$$\langle T(\boldsymbol{u}_j), \boldsymbol{u}_i \rangle = \sum_{k=1}^n A_j^k \langle \boldsymbol{u}_k, \boldsymbol{u}_i \rangle = \sum_{k=1}^n A_j^k \delta_{ik} = A_j^i.$$

Exchanging the roles of i and j gives $\langle \boldsymbol{u}_j, T(\boldsymbol{u}_i) \rangle = \langle T(\boldsymbol{u}_i), \boldsymbol{u}_j \rangle = A_i^j$.

The operator T is symmetric if and only if

$$A_j^i = \langle T(\boldsymbol{u}_j), \boldsymbol{u}_i \rangle = \langle \boldsymbol{u}_j, T(\boldsymbol{u}_i) \rangle = A_i^j \quad \text{for all } i \text{ and } j,$$

if and only if $[A_j^i] = [T]_{S'}^{S'}$ is a symmetric matrix. □

Remark 5.40. If $S' = (\boldsymbol{v}_j)_{j=1}^n$ is an arbitrary basis and T is symmetric, the expression

$$\sum_{k=1}^n A_j^k \langle \boldsymbol{v}_k, \boldsymbol{v}_i \rangle$$

is symmetric in i and j, but $[A_j^i] = [T]_{S'}^{S'}$ may not be symmetric.

Theorem 5.41. *Let A be a real, symmetric $n \times n$ matrix.*

(i) *Every eigenvalue of A is real.*

(ii) *Distinct eigenspaces of A are orthogonal.*

Proof. (i). Let $\lambda = \alpha + i\beta$ be an eigenvalue of A. It follows that the complex conjugate $\overline{\lambda} = \alpha - i\beta$ is also an eigenvalue of A. Consequently, the *real* matrix

$$\begin{aligned}\mathbf{A} &= (A - \lambda I_n)(A - \overline{\lambda} I_n) = A^2 - 2\alpha A + (\alpha^2 + \beta^2) I_n \\ &= (A - \alpha I_n)^2 + (\beta I_n)^2\end{aligned}$$

is not invertible. Since $A - \alpha I_n$ is symmetric, if $\boldsymbol{x}$ is a non-zero vector in the nullspace of $\mathbf{A}$, then

$$\begin{aligned}0 = \langle \mathbf{A}\boldsymbol{x}, \boldsymbol{x}\rangle &= \left\langle (A - \alpha I_n)^2 \boldsymbol{x}, \boldsymbol{x}\right\rangle + \left\langle (\beta I_n)^2 \boldsymbol{x}, \boldsymbol{x}\right\rangle \\ &= \langle (A - \alpha I_n)\boldsymbol{x}, (A - \alpha I_n)\boldsymbol{x}\rangle + \langle \beta\boldsymbol{x}, \beta\boldsymbol{x}\rangle \\ &= \|(A - \alpha I_n)\boldsymbol{x}\|^2 + \|\beta\boldsymbol{x}\|^2.\end{aligned}$$

By positive definiteness, $\beta = 0$ and $A\boldsymbol{x} = \alpha\boldsymbol{x}$.

(ii). If $\lambda^1 \neq \lambda^2$ are eigenvalues of A, with corresponding eigenvectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, then

$$\left\langle \lambda^1 \boldsymbol{u}_1, \boldsymbol{u}_2\right\rangle = \langle A\boldsymbol{u}_1, \boldsymbol{u}_2\rangle = \langle \boldsymbol{u}_1, A\boldsymbol{u}_2\rangle = \left\langle \boldsymbol{u}_1, \lambda^2 \boldsymbol{u}_2\right\rangle,$$

or $(\lambda^1 - \lambda^2)\langle \boldsymbol{u}_1, \boldsymbol{u}_2\rangle = 0$. Since $\lambda^1 - \lambda^2 \neq 0$, we have $\langle \boldsymbol{u}_1, \boldsymbol{u}_2\rangle = 0$. □

The Spectral Theorem

Theorem 5.42. *Let V be a subspace of $(\mathbf{R}^N, +, \cdot)$. If $T : V \to V$ is a symmetric linear operator, there exists an* orthonormal *T-eigenbasis of V.*

Proof. The proof proceeds by induction on the dimension of V. If $\dim V = 1$, then T is a scalar operator, and an arbitrary unit vector in V is an orthonormal eigenbasis.

Assume inductively for some $k \geq 1$ that the conclusion of the theorem holds for every subspace of dimension k, and let V be a subspace of dimension $(k + 1)$. Since T is symmetric, its eigenvalues are real. Pick an eigenvalue λ^{k+1} and a unit λ^{k+1}-eigenvector $\boldsymbol{u}_{k+1}$ in V, and let $V' = \operatorname{Span}(\boldsymbol{u}_{k+1})^\perp$ be the orthogonal complement of $\boldsymbol{u}_{k+1}$.

We claim T maps V' to V': A vector $\boldsymbol{y}$ is in V' if and only if $\langle \boldsymbol{u}_{k+1}, \boldsymbol{y}\rangle = 0$. Since T is symmetric and $\boldsymbol{u}_{k+1}$ is an eigenvector of T,

$$\langle \boldsymbol{u}_{k+1}, T(\boldsymbol{y})\rangle = \langle T(\boldsymbol{u}_{k+1}), \boldsymbol{y}\rangle = \langle \lambda \boldsymbol{u}_{k+1}, \boldsymbol{y}\rangle = 0;$$

that is, $T(\boldsymbol{y}) \in V'$. Since $\dim V' = k$, the inductive hypothesis guarantees there exists an orthonormal T-eigenbasis $(\boldsymbol{u}_j)_{j=1}^{k}$ of V'. Appending $\boldsymbol{u}_{k+1}$ gives an orthonormal T-eigenbasis $(\boldsymbol{u}_j)_{j=1}^{k+1}$ of V. □

Corollary 5.43. *If A is a symmetric $n \times n$ real matrix, there exists an orthogonal $n \times n$ matrix P such that $P^{\mathsf{T}}AP$ is diagonal.*

Theorem 5.42 has a geometric formulation:

Theorem 5.44. *Let V be a subspace of $(\mathbf{R}^N, +, \cdot)$. If $T : V \to V$ is a symmetric operator, $\lambda^1 > \lambda^2 > \cdots > \lambda^r$ are the eigenvalues of T, and $\Pi_j : V \to V$ is the orthogonal projection to the eigenspace E_{λ^j}, then:*

(i) $T = \sum_j \lambda^j \Pi_j$.

(ii) $I_V = \sum_j \Pi_j$.

Remark 5.45. (ii) recapitulates Theorem 3.25.

Example 5.46. (Symmetric rank-one operators on $\mathbf{R}^n$) Recall that a rank-one operator T on $\mathbf{R}^n$ has standard matrix $A = \boldsymbol{w}\boldsymbol{v}^{\mathsf{T}}$ for some non-zero vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbf{R}^n$.

The matrix A is symmetric if and only if

$$\boldsymbol{w}\boldsymbol{v}^{\mathsf{T}} = A = A^{\mathsf{T}} = \boldsymbol{v}\boldsymbol{w}^{\mathsf{T}},$$

if and only if $v^i w^j - v^j w^i = 0$ for all i, j, if and only if $\boldsymbol{v}$ and $\boldsymbol{w}$ are proportional. In this event, the eigenspaces are $E_0 = \mathrm{Span}(\boldsymbol{v})^{\perp}$, of dimension $(n-1)$, and $E_{\|\boldsymbol{v}\|^2} = \mathrm{Span}(\boldsymbol{v})$, of dimension 1.

In agreement with Theorem 5.44, T is $\|\boldsymbol{v}\|^2$ times orthogonal projection to $\mathrm{Span}(\boldsymbol{v})$.

5.5 Applications

For computing powers of a square matrix, diagonal matrices are formally simpler than general matrices. This section presents a selection of examples. Key properties are gathered here for reference; each is easily established with mathematical induction.

Proposition 5.47. *Let $n \geq 1$ be an integer. If A, B, and P and $n \times n$ matrices, with P invertible, then for all $m \geq 1$:*

(i) *If $A' = P^{-1}AP$, then $(A')^m = P^{-1}A^mP$.*

(ii) *If* $AB = BA$, *then* $(AB)^m = A^m B^m$.

(iii) *If* $(d^j)_{j=1}^n$ *are real numbers, and if* $A' = \mathrm{diag}[d^1, d^2, \dots, d^n]$, *then* $(A')^m = \mathrm{diag}\big[(d^1)^m, (d^2)^m, \dots, (d^n)^m\big]$.

The Fibonacci Numbers

The *Fibonacci sequence* is the real sequence $(x^m)_{m=1}^\infty$ defined by the *two-term recursion*

$$x^1 = x^2 = 1, \qquad x^{m+2} = x^m + x^{m+1} \quad \text{for } m \geq 1.$$

The first fifteen terms are 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610 .... An obvious inductive argument proves that each Fibonacci number is an integer. There is a remarkable closed formula, known to De Moivre but usually called "Binet's formula".

Theorem 5.48. *Let* $\lambda^+ = \frac{1}{2}(1+\sqrt{5})$ *and* $\lambda^- = \frac{1}{2}(1-\sqrt{5})$. *For* $m \geq 1$,

$$x^m = \frac{(\lambda^+)^m - (\lambda^-)^m}{\sqrt{5}} = \frac{(1+\sqrt{5})^m - (1-\sqrt{5})^m}{2^m\sqrt{5}}.$$

Proof. For $m \geq 1$, define the vector $\boldsymbol{x}^m = (x^m, x^{m+1})$ in $\mathbf{R}^2$. The Fibonacci recursion relation may be expressed in matrix form as

$$\boldsymbol{x}^{m+1} = \begin{bmatrix} x^{m+1} \\ x^{m+2} \end{bmatrix} = \begin{bmatrix} x^{m+1} \\ x^m + x^{m+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x^m \\ x^{m+1} \end{bmatrix} = A\boldsymbol{x}^m,$$

with A denoting the indicated 2×2 matrix. By induction on m,

$$\begin{bmatrix} x^m \\ x^{m+1} \end{bmatrix} = \boldsymbol{x}^m = A^{m-1}\boldsymbol{x}^1 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{m-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{for all } m \geq 1.$$

To calculate A^{m-1} in closed form, diagonalize A. The characteristic equation is

$$0 = \det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = \lambda^2 - \lambda - 1.$$

The quadratic formula gives the eigenvalues of A: $\lambda^+ = \frac{1}{2}(1+\sqrt{5})$ and $\lambda^- = \frac{1}{2}(1-\sqrt{5})$. The subsequent computation is done symbolically, using the identities

$$\lambda^+ + \lambda^- = 1, \quad \lambda^+ - \lambda^- = \sqrt{5}, \quad \lambda^-\lambda^+ = -1, \quad (\lambda^\pm)^2 = \lambda^\pm + 1.$$

Corresponding eigenvectors are

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ \lambda^+ \end{bmatrix}, \qquad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ \lambda^- \end{bmatrix}.$$

As guaranteed by Theorem 5.41, the eigenvalues of A are real and the corresponding eigenvectors are orthogonal.

Let $S = (\boldsymbol{e}_j)_{j=1}^2$ be the standard basis of $\mathbf{R}^2$ and $S' = (\boldsymbol{v}_j)_{j=1}^2$ the A-eigenbasis. The transition matrix $P = [I_2]_{S'}^S$ and its inverse are

$$P = \begin{bmatrix} 1 & 1 \\ \lambda^+ & \lambda^- \end{bmatrix}, \qquad P^{-1} = \frac{1}{\lambda^+ - \lambda^-}\begin{bmatrix} -\lambda^- & 1 \\ \lambda^+ & -1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} -\lambda^- & 1 \\ \lambda^+ & -1 \end{bmatrix}.$$

(Note that $P^{-1} \neq P^{\mathsf{T}}$; the columns of P are not of unit length.)

If $T : \mathbf{R}^2 \to \mathbf{R}^2$ is the operator whose standard matrix is A, then

$$A' = \begin{bmatrix} \lambda^+ & 0 \\ 0 & \lambda^- \end{bmatrix} = [T]_{S'}^{S'} = [I_2]_S^{S'}[T]_S^S[I_2]_{S'}^S = P^{-1}AP.$$

Writing $A = PA'P^{-1} = \operatorname{diag}[\lambda^+, \lambda^-]$ and remembering $\lambda^+\lambda^- = -1$,

$$\begin{aligned} A^{m-1} &= P(A')^{m-1}P^{-1} \\ &= \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 1 \\ \lambda^+ & \lambda^- \end{bmatrix}\begin{bmatrix} (\lambda^+)^{m-1} & 0 \\ 0 & (\lambda^-)^{m-1} \end{bmatrix}\begin{bmatrix} -\lambda^- & 1 \\ \lambda^+ & -1 \end{bmatrix} \\ &= \frac{1}{\sqrt{5}}\begin{bmatrix} (\lambda^+)^{m-2} - (\lambda^-)^{m-2} & (\lambda^+)^{m-1} - (\lambda^-)^{m-1} \\ (\lambda^+)^{m-1} - (\lambda^-)^{m-1} & (\lambda^+)^m - (\lambda^-)^m \end{bmatrix}. \end{aligned}$$

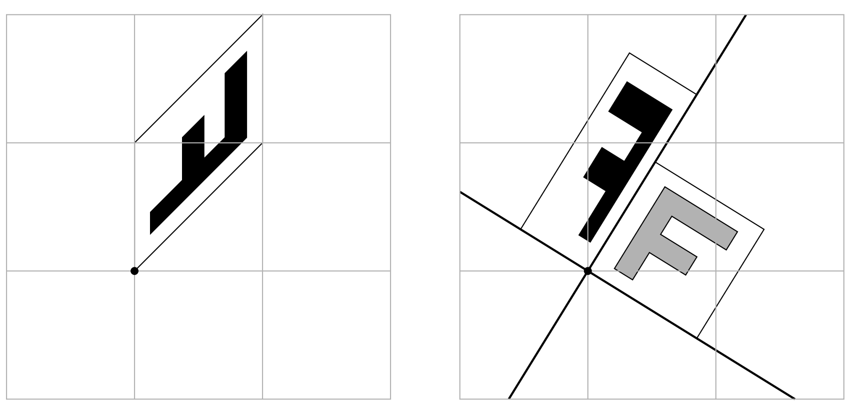


Figure 5.1: The action of T on S, and on an orthonormal eigenbasis.

Consequently,

$$\begin{aligned}
\begin{bmatrix} x^m \\ x^{m+1} \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{m-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&= \frac{1}{\sqrt{5}} \begin{bmatrix} (\lambda^+)^{m-2} - (\lambda^-)^{m-2} & (\lambda^+)^{m-1} - (\lambda^-)^{m-1} \\ (\lambda^+)^{m-1} - (\lambda^-)^{m-1} & (\lambda^+)^{m} - (\lambda^-)^{m} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&= \frac{1}{\sqrt{5}} \begin{bmatrix} \big((\lambda^+)^{m-2} + (\lambda^+)^{m-1}\big) - \big((\lambda^-)^{m-2} + (\lambda^-)^{m-1}\big) \\ \big((\lambda^+)^{m-1} + (\lambda^+)^{m}\big) - \big((\lambda^-)^{m-1} + (\lambda^-)^{m}\big) \end{bmatrix}.
\end{aligned}$$

In particular, since $1 + \lambda^\pm = (\lambda^\pm)^2$,

$$\begin{aligned}
x^m &= \frac{\big((\lambda^+)^{m-2} + (\lambda^+)^{m-1}\big) - \big((\lambda^-)^{m-2} + (\lambda^-)^{m-1}\big)}{\sqrt{5}} \\
&= \frac{(\lambda^+)^m - (\lambda^-)^m}{\sqrt{5}}.
\end{aligned}$$
□

Remark 5.49. Each power in the numerator may be expanded using the binomial theorem. It's a good exercise to simplify the numerator in this way, and to prove the formula really does generate only integers.

Matrix Exponentiation

Definition 5.50. The *trace* of an $n \times n$ matrix A is the sum of its diagonal entries,

$$\operatorname{tr}(A) = A_1^1 + \cdots + A_n^n = \sum_{j=1}^n A_j^j.$$

Proposition 5.51. *Let A and B be $n \times n$ matrices.*

(i) $\operatorname{tr}(BA) = \operatorname{tr}(AB)$.

(ii) *If P is invertible, then* $\operatorname{tr}(P^{-1}AP) = \operatorname{tr}(A)$.

(iii) *If A is diagonalizable, then* $\operatorname{tr}(A)$ *is the sum of the eigenvalues, counting multiplicity.*

Proof. (i). Since $(AB)_j^i = \sum_k A_k^i B_j^k$,

$$\operatorname{tr}(AB) = \sum_{i=1}^n \sum_{k=1}^n A_k^i B_i^k = \sum_{k=1}^n \sum_{i=1}^n B_i^k A_k^i = \operatorname{tr}(BA).$$

(ii). By (i), $\operatorname{tr}\big(P^{-1}(AP)\big) = \operatorname{tr}\big((AP)P^{-1}\big) = \operatorname{tr}(A)$.

(iii). Suppose $A' = P^{-1}AP$. By (ii) $\operatorname{tr}(A) = \operatorname{tr}(A')$.

Since A' is diagonal, the eigenvalues of A' are its diagonal entries, so $\operatorname{tr}(A')$ is the sum of the eigenvalues, say $\Lambda_{A'}$.

The matrices A and A' have the same characteristic polynomial, hence the same eigenvalues. In particular, $\Lambda_{A'} = \Lambda_A$, the sum of the eigenvalues of A.

Combining these observations, $\operatorname{tr}(A) = \operatorname{tr}(A') = \Lambda_{A'} = \Lambda_A$. □

Definition 5.52. Let $A = [A^i_j]$ be an $n \times n$ real (or complex) matrix. The *exponential series* is defined by

$$\exp(tA) = \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} = I + tA + t^2\,\frac{A^2}{2!} + t^3\,\frac{A^3}{3!} + \dots.$$

Remark 5.53. Since there are only finitely many entries of A, there is a real number M such that $|A^i_j| \leq M$ for all i and j. By induction on m, $|(A^m)^i_j| \leq (Mn)^k$ for all $k \geq 0$; for example, if $m = 1$, we have

$$|(A^2)^i_j| = \left|\sum_{\ell=1}^{n} A^i_\ell A^\ell_j\right| \leq \sum_{\ell=1}^{n} |A^i_\ell A^\ell_j| \leq M^2 n \leq (Mn)^2.$$

It follows that each entry in the exponential series is bounded in absolute value by the convergent series

$$\sum_{k=0}^{\infty} \frac{(tMn)^k}{k!} = e^{tMn}.$$

Example 5.54. If $A = \operatorname{diag}[d^1, d^2, \dots, d^n]$, then A^m is the diagonal matrix whose entries are the mth powers of the entries of A, so

$$\exp(tA) = \operatorname{diag}[e^{td^1}, e^{td^2}, \dots, e^{td^n}].$$

Example 5.55. If N is the 4×4 nilpotent block, its fourth power is $\mathbf{0}^{4\times 4}$, so N, N^2, N^3, and $\exp(tN)$ are

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} \\ 0 & 1 & t & \frac{t^2}{2!} \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Example 5.56. If $N = B_n(0)$ is the $n \times n$ nilpotent block, then N^m has a diagonal of 1's m rows above the diagonal, and $N^n = \mathbf{0}^{n\times n}$. The exponential series is a polynomial, and

$$\exp(tN) = \begin{bmatrix} 1 & t & \frac{t^2}{2!} & \dots & \frac{t^{n-2}}{(n-2)!} & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \dots & \frac{t^{n-3}}{(n-3)!} & \frac{t^{n-2}}{(n-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & t & \frac{t^2}{2!} \\ 0 & 0 & 0 & \dots & 1 & t \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Theorem 5.57. *Let A and B be $n \times n$ matrices.*

(i) *If P is an invertible $n \times n$ matrix, then*
$$\exp(tP^{-1}AP) = P^{-1}\exp(tA)P.$$

(ii) *If $BA = AB$, then* $\exp\bigl(t(A+B)\bigr) = \exp(tA)\exp(tB)$.

(iii) *If A is diagonalizable, then* $\det\exp(tA) = e^{t\operatorname{tr}(A)}$.

(iv) $\exp(tA^{\mathsf{T}}) = \bigl[\exp(tA)\bigr]^{\mathsf{T}}$.

(v) *If $A^{\mathsf{T}} = -A$, then $\exp(tA)$ is orthogonal and has determinant* 1.

Proof. (i). This is a consequence of the identity $(P^{-1}AP)^k = P^{-1}A^kP$:
$$\exp(tP^{-1}AP) = \sum_{k=0}^{\infty}\frac{(tP^{-1}AP)^k}{k!} = \sum_{k=0}^{\infty}P^{-1}\frac{(tA)^k}{k!}P = P^{-1}\exp(tA)P.$$

(ii). This follows from the binomial formula
$$\frac{(A+B)^k}{k!} = \sum_{i=0}^{k}\frac{A^i}{i!}\,\frac{B^{k-i}}{(k-i)!} = \sum_{i+j=k}\frac{A^i}{i!}\,\frac{B^j}{j!},$$
which holds for commuting matrices, together with the "Cauchy product formula" for absolutely convergent double series:
$$\begin{aligned}\exp\bigl(t(A+B)\bigr) &= \sum_{k=0}^{\infty}\frac{\bigl(t(A+B)\bigr)^k}{k!} = \sum_{k=0}^{\infty}\sum_{i+j=k}\frac{(tA)^i}{i!}\,\frac{(tB)^j}{j!} \\ &= \left(\sum_{i=0}^{\infty}\frac{(tA)^i}{i!}\right)\left(\sum_{j=0}^{\infty}\frac{(tB)^j}{j!}\right) = \exp(tA)\exp(tB).\end{aligned}$$

(iii). Suppose $P^{-1}AP = A' = \operatorname{diag}[d^1, \dots, d^n]$ for some invertible matrix P. By (i) and Example 5.54,

$$P^{-1} \exp(tA) P = \exp(tA') = \operatorname{diag}[e^{td^1}, e^{td^2}, \dots, e^{td^n}].$$

Taking determinants,

$$\det \exp(tA) = \det \exp(tA') = \prod_{k=1}^{n} e^{td^k} = e^{t \sum_k d^k} = e^{t \operatorname{tr}(A')} = e^{t \operatorname{tr}(A)}.$$

(iv). This is an immediate consequence of $(A^k)^{\mathsf{T}} = (A^{\mathsf{T}})^k$.

(v). Clearly $\exp(\mathbf{0}^{n \times n}) = I_n$, so by (ii), $\exp(-tA) = \exp(tA)^{-1}$. If $A^{\mathsf{T}} = -A$. then

$$\exp(tA)^{-1} = \exp(-tA) = \exp(tA^{\mathsf{T}}) = \exp(tA)^{\mathsf{T}},$$

which proves $\exp(tA)$ is orthogonal.

Since the diagonal entries of a skew-symmetric matrix are all 0, $\operatorname{tr}(A) = 0$. By (iii), $\det \exp(tA) = e^{t \operatorname{tr}(A)} = e^0 = 1$. □

Remark 5.58. An orthogonal $n \times n$ matrix of determinant 1 is, by definition, a *Euclidean rotation*. Part (v) of the theorem says that a skew-symmetric matrix is an "infinitesimal" rotation.

Example 5.59. If

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \text{then } \exp(tA) = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix}.$$

To prove this, note that $A^2 = -I_2$, from which it follows immediately that $A^{2k} = (-1)^k I_2$ and $A^{2k+1} = (-1)^k A$. Splitting the exponential series into terms of even and odd degree and recalling the power series for the circular functions,

$$\cos t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!}, \qquad \sin t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!},$$

we have

$$\begin{aligned} \exp(tA) &= \sum_{k=0}^{\infty} \frac{(tA)^{2k}}{(2k)!} + \sum_{k=0}^{\infty} \frac{(tA)^{2k+1}}{(2k+1)!} \\ &= \left(\sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!} \right) I_2 + \left(\sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!} \right) A \\ &= (\cos t) I_2 + (\sin t) A. \end{aligned}$$

Exercises

Exercise 5.1. Find the eigenvalues and bases for the eigenspaces of the following matrices, and determine which are diagonalizable:

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 1 & 0 & 0 & 4 \end{bmatrix} \qquad \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix} \qquad \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \qquad \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Exercise 5.2. On the vector space $V = \mathcal{F}(\mathbf{R}, \mathbf{R})$ of all real-valued functions on $\mathbf{R}$ under function addition and scalar multiplication, let R be the "domain reflection" operator defined by $(Rf)(t) = f(-t)$.

(a) Show R is a linear involution.

(b) Identify the eigenspaces of R, and show that every function can be written uniquely as a sum of eigenvectors.

(c) Find the decomposition of $f(t) = e^t$ into eigenvectors.

Exercise 5.3. Prove Proposition 5.37.

Exercise 5.4. An operator $\Pi : V \to V$ is a *projection* if $\Pi^2 = \Pi$.

(a) Prove that if Π is a projection, then $I_V - \Pi$ is a projection, and the only eigenvalues of Π are 0 and 1.

(b) Assuming "optimistically" that every vector $\boldsymbol{v}$ in V decomposes as a sum $\boldsymbol{v}^0 + \boldsymbol{v}^1$ of eigenvectors, find formulas for $\boldsymbol{v}^0$ and $\boldsymbol{v}^1$ in terms of $\boldsymbol{v}$ and $\Pi(\boldsymbol{v})$.

(c) Show that the vectors $\boldsymbol{v}^0$ and $\boldsymbol{v}^1$ found in part (b) are eigenvectors of Π. Conclude that Π is diagonalizable.

(d) Find a projection on $\mathbf{R}^2$ whose standard matrix is *not* symmetric.

Exercise 5.5. If $\Pi : V \to V$ is a projection (preceding exercise), the operator $R = I_v - 2\Pi$ is a *reflection.*

(a) Prove that a reflection is an involution.

(b) How are the eigenvalues and eigenspaces of Π related to the eigenspaces of R?

(c) Find a reflection on $\mathbf{R}^2$ whose standard matrix is *not* symmetric.

Exercise 5.6. Let $T : V \to V$ be an operator satisfying $T^3 = T$.

(a) Find the possible eigenvalues of T.

(b) Assuming "optimistically" that every vector $\boldsymbol{v}$ in V decomposes as a sum of eigenvectors, find formulas for the each summand in terms of $\boldsymbol{v}$, $T(\boldsymbol{v})$ and $T^2(\boldsymbol{v})$.

(c) Show that the vectors found in part (b) are eigenvectors of T. Conclude that T is diagonalizable.

Exercise 5.7. Let A be a non-zero $n \times n$ matrix satisfying $A^k = \mathbf{0}^{n\times n}$ for some integer $2 \leq k \leq n$.

(a) Prove that every eigenvalue of A is 0.

(b) If $A^3 = \mathbf{0}^{n\times n}$, prove $(I + A)^{-1} = I - A + A^2$.

(c) Find a formula for $(I + A)^{-1}$ if $A^k = \mathbf{0}^{n\times n}$ for some $k > 3$.

(d) Prove $I + A$ is not diagonalizable.
Hint: What are the eigenvalues?

Exercise 5.8. Let $S = (\boldsymbol{v}_j)_{j=1}^n$ be a basis of $\mathbf{R}^n$. Prove S is orthonormal if and only if $\boldsymbol{v}_1\boldsymbol{v}_1^\mathsf{T} + \cdots + \boldsymbol{v}_n\boldsymbol{v}_n^\mathsf{T} = I_n$.

Exercise 5.9. Let B be an arbitrary $n \times n$ real matrix, and put $A = B^\mathsf{T} B$, noting that A is symmetric. Prove that every eigenvalue of A is non-negative, and 0 is an eigenvalue of A if and only if B is not invertible.
Hint: First show that $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \|B\boldsymbol{x}\|^2$ for every $\boldsymbol{x}$ in $\mathbf{R}^n$.

Exercise 5.10. Let a, b, and c be real numbers.

(a) Find the eigenvalues and eigenspaces of $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$.

(b) Prove (by direct computation) that the eigenspace of A are orthogonal.

(c) Prove the eigenvalues have opposite sign if and only if $ac - b^2 < 0$.

(d) Prove the eigenvalues have the same sign if and only if $ac - b^2 > 0$, and their mutual sign is the sign of a.

Exercise 5.11. Let $A = \begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix}$. Find a formula for A^n, $n > 0$ an arbitrary integer.

Exercise 5.12. Show that if A is skew-symmetric and λ is a real eigenvalue, then $\lambda = 0$.
Hint: Let $\boldsymbol{v}$ be a λ-eigenvector, and compute $\langle A\boldsymbol{v}, \boldsymbol{v}\rangle$.

Exercise 5.13. Let A and B be $n \times n$ matrices. We say A and B are *simultaneously diagonalizable* if there exists an invertible matrix P such that $A' = P^{-1}AP$ and $B' = P^{-1}BP$ are diagonal.

Prove that simultaneously diagonalizable matrices commute, i.e., $BA = AB$.
Hint: What can you say about the product of diagonal matrices?

Exercise 5.14. Let A and B be *commuting* $n \times n$ matrices.

(a) Show that if $\boldsymbol{v}$ is a λ-eigenvector of A, then $B\boldsymbol{v}$ is in the λ-eigenspace of A. That is, B preserves the eigenspaces of A.

(b) Prove that if A is diagonalizable and B is invertible, then A and B are simultaneously diagonalizable.
Hint: Show that A and B^{-1} commute, then use part (a) to show the eigenspaces of B are precisely the eigenspaces of A.

(c) Remove the hypothesis in part (b) that B is invertible.
Suggestion: B is diagonalized by P if and only if $B + cI_n$ is; show that there exists a real number c so that $B + cI_n$ is invertible.

Index